University of Zagreb

FACULTY OF SCIENCE

DEPARTMENT OF BIOLOGY

Vanja Tadić

CONSTRUCTION OF MOLECULAR TOOLS BASED ON CRISPR/Cas9 SYSTEM FOR EPIGENETIC MODULATION OF GENE EXPRESSION

DOCTORAL THESIS

Zagreb, 2019.

Sveučilište u Zagrebu

PRIRODOSLOVNO-MATEMATIČKI FAKULTET

BIOLOŠKI ODSJEK

Vanja Tadić

KONSTRUKCIJA MOLEKULARNIH ALATA TEMELJENIH NA SUSTAVU CRISPR/Cas9 ZA EPIGENETIČKU MODULACIJU EKSPRESIJE GENA

DOKTORSKI RAD

Zagreb, 2019.

Ovaj doktorski rad izrađen je na Zavodu za molekularnu biologiju, Biološkog odsjeka Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu, pod voditeljstvom prof. dr. sc. Vlatke Zoldoš, u sklopu Sveučilišnog poslijediplomskog doktorskog studija Biologije, pri Biološkom odsjeku Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu.

Vlatka Zoldoš, PhD, is a professor of Genetics and Epigenetics at the University of Zagreb Faculty of Science. She obtained her PhD in 2000 from the University Paris Sud XI, France and the University of Zagreb, Croatia, in the frame of dual mentorship program. Professor Zoldoš was working for 15 years in the field of plant biology where she achieved considerable recognition of the international scientific community. In 2008, she established Laboratory for epigenetics at the Department of molecular biology and shifted her main research interest to epigenetics of protein glycosylation associated with human diseases. Her pioneering efforts in this field resulted in several prominent papers including the first demonstration of epigenetic regulation of plasma protein glycosylation as well as glycosylation of a single protein. She was also the first who showed reversibility of the cell N-glycome changes following epigenetic modulation using epigenetic inhibitors, and she was the first who developed CRISPR/Cas9-based tool for targeted DNA methylation.

She published over 40 scientific papers in the international scientific journals with high impact factor such as Nature Communications, Nature Reviews Gastroenterology and Hepatology, Nucleic Acids Research, Current Opinion in Chemical Biology and Epigenetics. She is involved in several large consortia and projects funded by European foundations (such as FP7 and Horizon2020 programs), several national projects, and her group is a part of the Centre of Scientific Excellence in Personalized Healthcare. She is a reviewer and associated editor in several international scientific journals with high impact factor.

ZAHVALE

Hvala mentorici, prof. dr. sc. Vlatki Zoldoš na ukazanom povjerenju i prilici da budem član istraživačke grupe Laboratorija za epigenetiku. Zahvalna sam na svim pruženim prilikama za učenje na kongresima i radionicama, kao i na savjetima i konstruktivnim kritikama tijekom trajanja projekta i izrade ovog doktorata.

Hvala kolegama iz grupe i članovima Zavoda za molekularnu biologiju koji su mi znali nesebično pomoći kada je to bilo potrebno.

Paula i Marija, hvala vam na svemu što ste me naučile o radu u laboratoriju, radu sa studentima i o životu, bila mi je čast raditi s vama.

Hvala kolegi Martinu na slušanju, strpljenju i svim diskusijama koje su mi pomogle tijekom izrade ovog doktorata.

Ivona, Dora, Renata, Luka i Melanija, hvala vam što ste bili divni kolege i prijatelji.

Hvala izv. prof. dr. sc. Petri Korać na svim savjetima, dobrom raspoloženju i diskusijama zbog kojih sam postajala bolja znanstvenica i osoba.

Zahvaljujem svojoj obitelji na bezuvjetnoj podršci i ljubavi, bez vas za mene apsolutno ništa ne bi imalo smisla.

University of Zagreb

Faculty of Science

Department of Biology

Doctoral thesis

CONSTRUCTION OF MOLECULAR TOOLS BASED ON CRISPR/Cas9 SYSTEM FOR EPIGENETIC MODULATION OF GENE EXPRESSION

VANJA TADIĆ

Faculty of Science, University of Zagreb

Repurposing of platforms for targeted genome engineering, zinc finger nucleases (ZFNs), transcription activator-like nucleases (TALENs) and especially CRISPR/Cas9 system, has enabled targeted manipulation of epigenetic marks and new insights into epigenetic regulatory mechanisms. In this work, I have fused the catalytic domain of human TET1 (Ten-Eleven Translocation 1) protein C-terminally to the catalytically inactive Cas9 (dCas9) from *Streptococcus pyogenes* to create the dCas9-TET1 molecular tool. I have also tested the activity of N-terminal fusion of TET1-dCas9. The TET1 protein is a methylcytosine dioxygenase involved in active demethylation of 5-methylctosines to 5-hydroxymethylcytosines. In context of gene expression, methylated CpG dinucleotides within gene promoters are usually associated with gene silencing, and unmethylated CpGs with active gene transcription. Constructed molecular tools were successfully used for targeted CpG demethylation in promoter regions of two candidate genes, *MGAT3* and *LAMB1*. I also showed that following demethylation the *MGAT3* gene changed expression level. I have compared *MGAT3* gene activation rates achieved following demetlylation using dCas9-TET1 with a more conservative approach of VPR-dCas9 mediated direct activation. For this approach, I used an N-terminal fusion of VPR activation domain and dCas9 from *Staphylococcus aureus* (dSaCas9).

(120 pages, 30 figures, 5 tables, 169 references, original in English)

Keywords: CRISPR/dCas9 system, epigenetic editing, TET1, DNA methylation, gene expression

Supervisor: Professor Vlatka Zoldoš, PhD

Reviewers: Assistant Professor Aleksandar Vojta, PhD

Professor Srećko Gajović, PhD

Melita Vidaković, PhD, Scientific Adviser

Sveučilište u Zagrebu Doktorski rad

Prirodoslovno-matematički fakultet

Biološki odsjek

KONSTRUKCIJA MOLEKULARNIH ALATA TEMELJENIH NA SUSTAVU CRISPR/Cas9 ZA EPIGENETIČKU MODULACIJU EKSPRESIJE GENA

VANJA TADIĆ

Prirodoslovno-matematički fakultet Sveučilišta u Zagrebu

Prenamjene alata za genetičko inženjerstvo, baziranih na *zinc finger* nukleazama (ZFN), *transcription activator-like effector* nukleazama (TALEN) te posebno sustavu CRISPR/Cas9, pružile su mogućnost ciljanog modificiranja epigenoma i time razjašnjavanje uloge epigenetičkih modifikacija u regulaciji ekspresije gena. U ovom radu napravila sam C-terminalnu fuziju katalitičke domene proteina TET1 (Ten-Eleven Translocation 1) i katalitički inaktivne nukleaze Cas9 (dCas9) izolirane iz vrste *Streptococcus pyogenes* i konstruirala dCas9-TET1 molekularni alat za ciljanu demetilaciju DNA. Testirala sam i aktivnost N-terminalne fuzije TET1-dCas9. Protein TET1 katalizira pretvorbu 5-metilcitozina u 5-hidroksimetilcitozin, što predstavlja prvi korak u aktivnoj demetilaciji DNA. Korelacijske analize povezuju hipermetilirane CpG dinukleotide u promotorima gena s represijom transkripcije, a hipometilirane CpG dinukleotide s aktivnom transkripcijom. Konstruirani alati uspješno su upotrijebljeni za ciljanu demetilaciju CpG dinukleotida unutar promotora kandidat gena *MGAT3* i *LAMB1*. Pokazala sam da je nakon demetilacije ciljanih CpG mjesta u promotoru gena *MGAT3* porasla razina transkripata ovoga gena. Usporedila sam efekt aktivacije gena *MGAT3* postignut ciljanom demetilacijom dCas9-TET1 alatom s konzervativnijim pristupom direktne aktivacije pomoću VPR-dCas9 alata. U ovu svrhu upotrijebila sam N-terminalnu fuziju aktivacijske domene VPR i dCas9 iz vrste *Staphylococcus aureus* (dSaCas9).

(120 stranica, 30 slika, 5 tablica, 169 literaturnih navoda, jezik izvornika engleski)

Ključne riječi: sustav CRISPR/Cas9, epigenetičko inženjerstvo, TET1, metilacija DNA, ekspresija gena

Mentor: prof. dr. sc. Vlatka Zoldoš

Ocjenjivači: doc. dr. sc. Aleksandar Vojta

prof. dr. sc. Srećko Gajović

dr.sc. Melita Vidaković, znanstveni savjetnik

Table of contents

1. Introduction

Regulation of gene transcription is a complex process controlled at several hierarchical levels. Transcription occurs upon binding of transcription factors (TFs) which recruit RNA polymerase to the gene promoter region, but many complexes of different proteins are also involved in the process, needed for optimal chromatin remodelling *via* recruitment of various components of the large epigenetic machinery and optimal DNA folding and positioning (1). Maintenance of specific gene expression patterns defines cell identity and function. Epigenetic mechanisms, encompassing DNA methylation, posttranslational histone modifications, chromatin remodelling, positional information, action of non-coding RNAs and histone variants regulate and maintain gene expression patterns through cell divisions and development (2–4).

Epigenetic mechanisms have been recognized as crucial mediators of gene expression related to developmental biology (5), nevertheless only recently along with development of engineered molecular tools for targeted epigenome editing the straightforward approaches are available for unravelling these mechanisms in more detail. This was achieved by repurposing of platforms for targeted genome engineering for targeted manipulation of epigenetic marks. Most common strategies in construction of molecular tools for genome editing to date have been based on programmable nucleases created with Zinc finger (ZF) and Transcription activator-like effector (TALE) proteins. These systems possess the required ability to target and cleave a specific DNA sequence, but the assembly is quite tedious and tools have significant off-target activity (6–9). A novel technology based on CRISPR/Cas9 (clustered regularly interspaced palindromic repeats/CRISPR associated protein 9) system has recently emerged offering a potential for (epi)genome editing due to its programming simplicity (10,11).

Repurposed CRISPR/Cas9 systems, using the inactivated Cas9 nuclease (dCas9) for targeting by small single guide RNA (sgRNA) molecule, have been successfully used for delivery of various effector domains to any sequence of interest, acting as epigenome editing tools. Several dCas9-based tools have been developed for targeted gene activation by recruitment of activation domains in mechanism called CRISPR activation (CRISPRa) and for targeted gene repression by steric interference or recruitment of repressor domains in so called

CRISPR interference (CRISPRi) (12). Fusions of dCas9 and catalytic domains of various epigenetic writers (enzymes that put on epigenetic marks) and erasers (enzymes that put off epigenetic marks) have also been developed for more specific epigenetic manipulations. Fusion of dCas9 and the catalytic core of the human acetyltransferase p300 successfully acetylyzed histone H3 lysine 27 (H3K27) at targeted sites and activated gene expression from promoters and enhancers (13). Through previous efforts of my group, we successfully repurposed the CRISPR/Cas9 system for targeted DNA methylation by fusing dCas9 with a catalytic domain of human DNA methyltransferase 3A (DNMT3A) and achieved gene repression through targeted promoter CpG methylation (14).

Hypothesis of this doctoral thesis is that demethylation of certain CpG sites regulates gene expression. Therefore, the aim was construction of CRISPR/dCas9 molecular tool for targeted cytosine demethylation, which would enable interrogation of the functional relevance of individual CpG sites within promoters of the candidate genes for transcriptional regulation.

DNA methylation is a process in which a methyl group (-CH_3) is added to the fifth carbon atom of cytosine generating 5-methylcytosine (5mC) base. In mammals, more than 98% of DNA methylation occurs in CpG dinucleotides in somatic cells. It is an important epigenetic mechanism regulating gene expression, embryonic development, genomic imprinting, maintenance of genome stability and chromatin structure (3). Establishment and maintenance of genome methylation pattern plays an important role in normal cell function. Aberrant methylation status leads to development of various multifactorial diseases, including inflammation, diabetes and cancer (15). Protein glycosylation is believed to be epigenetically regulated and it could be an important source of variability in a population and act as adaptory mechanism (16), while epigenetic deregulation of glyco-genes results in aberrant protein glycosylation and can lead to disease (17,18). Immunoglobulin G (IgG) is known to have different properties depending on its N-glycome, changing the immune response it mediates (19). Our group is primarily interested in epigenetic regulation of IgG glycosylation in inflammatory diseases and epigenetic deregulation in diabetes. In the context of gene expression, methylated CpG dinucleotides within gene promoters and other regulatory regions, such as enhancers, are usually associated with gene silencing, and unmethylated CpGs with gene transcription (3,20). Gene repression is achieved either through the interference of methyl group with TF binding or through recruitment of methyl-CpG binding

domain (MBD) proteins, which are part of repressive transcription complex consisting of histone deacetylases (HDACs) and histone methyltransferases (HTMs), responsible for repressive histone modifications, as well as of chromatin remodelers involved in formation of closed chromatin conformation (4). CpG demethylation associated with transcription activation can occur by active 5mC oxidation to 5-hydroxymethylcytosine (5hmC) catalysed by the TET enzymes or by replication-dependent passive CpG demethylation promoted by TF binding (21,22).

General goal was to develop CRISPR/dCas9-based molecular toolbox which would be modular and easily reconfigured for a range of applications in direct gene regulation and epigenome editing. Primary focus was on construction and validation of CRISPR/Cas9-TET1 molecular tool for targeted cytosine demethylation.

Specific goals were:

i. Construction of expression cassette comprising Cas9 protein fused to functional domain targeted to specific genomic loci, with desired selection marker (fluorescent protein or antibiotic resistance). Functional domains chosen for delivery will be domains for activation, repression and DNA methylation and demethylation. Promoter choice will allow for inducible expression.

ii. Construction of plasmid vector for expression of TET1 catalytic domain fused with dCas9 nuclease.

iii. Choice of DNA regions in genes of interest suitable for validation of constructed molecular tools.

iv. Vector delivery to HEK293 and HEK293 FreeStyle cell lines, and optionally to B-lymphocytes.

v. Determining the methylation status of targeted DNA regions with bisulfite pyrosequencing tests designed for methylation status analysis of single, specific CpG sites.

vi. Analysis of hydroxymethylation status of targeted DNA region with bisulfite pyrosequencing tests designed for hydroxymethylation analysis.

vii. Functional analysis (of gene expression levels) after epigenetic modulation with constructed molecular tool.

Here I have constructed a plasmid vector for expression of the TET1 catalytic domain fused with dCas9 nuclease from *Streptococcus pyogenes* for targeted DNA demethylation. This construct catalyses oxidation of 5-methylcytosines (5mCs), creating 5-hydroxymethylcytosines (5hmCs) which are then turned to unmethylated cytosines via subsequent oxidation and repair. For purpose of the negative control, the catalytically inactive TET1 was generated by site-directed mutagenesis of the active site motif and inserted into CRISPR/dCas9 expression vector. Selection markers PuroR and EGFP were inserted downstream of T2A peptide linker at the C-terminus of the dCas9-TET1 fusion to allow for either puromycin selection or GFP screening of transfected cells. For delivery of the dCas9-TET1 tool to specific CpG sites within promoters of the candidate genes, *MGAT3* and *LAMB1*, I have designed several specific sgRNAs. Regions within promoters of the *MGAT3* and *LAMB1* genes which are highly methylated in HEK293 cells were selected based on the previous pyrosequencing results. Designed sgRNAs were synthesized by oligo annealing and subsequently inserted in the vector downstream of the human U6 promoter. The genes *MGAT3* and *LAMB1* were chosen as candidate genes because both genes are GWAS hits for IgG glycosylation and inflammatory bowel disease (IBD) (23). Direct manipulation of *MGAT3* and *LAMB1* using CRISPR/dCas9-based tools would contribute to understanding of mechanisms involved in regulation of IgG glycosylation, and possibly revealing how aberrant IgG glycosylation contributes to inflammation.

HEK293 cell line was used as a basic model in research and development of method for targeted DNA demethylation. I also tried to use the CRISPR/dCas9-based molecular tools in suspension cell line HEK293 FreeStyle transient system (24), which is an excellent biologically relevant model cell line because it secretes IgG molecules. I also tried to import the CRISPR/dCas9 constructs to primary B-lymphocyte culture by lentiviral delivery.

However, the optimization of the delivery method to HEK293 FreeStyle and primary B-cells turned out to be an incredibly tedious task and the scope of these experiments surpassed the extent of this research.

DNA from crude cell lysates of dCas9-TET1 transfected cells was bisulfite converted and purified for methylation analysis at targeted regions, and for analysis of hydroxymethylation status. CpG methylation status was determined using pyrosequencing after bisulfite conversion. However, conventional bisulfite sequencing methods are not precise in base definition when it comes to 5mC and 5hmC. Bisulfite treatment causes deamination of unmethylated cytosines to uracils, while both 5mCs and 5hmCs remain resistant to deamination during treatment and are therefore both sequenced as cytosines (25). Therefore, the method was not sensitive enough for determination of hydroxymethylation status, for which an oxidative bisulfite conversion approach was used followed by bisulfite pyrosequencing tests for targeted genes. Activity profile of dCas9-TET1 molecular tool was analysed on *MGAT3* locus, and I have also tested the activity of N-terminal fusion of TET1 catalytic domain and dCas9 from *Streptococcus pyogenes* (TET1-dCas9) on the same locus.

For gene expression analyses using quantitative real-time PCR (RT-qPCR), total RNA was isolated from transfected cells at the same time as DNA. Changes in gene expression were determined using ddCt relative quantification method. I have also compared *MGAT3* gene activation rates achieved following demethylation using dCas9-TET1 with a more conservative approach of VPR-dCas9 mediated direct activation. VPR is a chimeric activation domain composed of the activation domains of VP64, p65 and Rta (26). I used a construct containing an N-terminal fusion of catalytically inactive Cas9 from *Staphylococcus aureus* (SaCas9) and VPR activation domain. The VPR-dCas9 construct was targeted to specific region in the *MGAT3* gene using specific sgRNAs.

2. Literature review

2.1 Genome editing

With discovery of restriction endonucleases in the 1970s (27,28), first advances were made in the so called recombinant DNA technology, which gradually evolved to a point were targeted genome editing is broadly used for modifications of virtually any targeted DNA sequence in living cells and organisms. Targeted genome engineering is the modification of the genome at a precisely chosen locus. Selectively inactivating specific genes and measuring following effects on a phenotype could provide a direct insight into gene functions (29). The simplest form of genome editing is gene knockout by introducing a double stranded break (DSB) to targeted DNA sequence which results in recruitment of endogenous repair mechanisms of the cell. DSBs are repaired by homology-directed repair (HDR) or non-homologous end joining (NHEJ). HDR requires a homologous DNA template to guide the high fidelity repair and can be used for insertion of specific sequence flanked with homology arms at the target site resulting in gene correction (30,31), reporter insertion, etc., while NHEJ happens randomly and is error-prone, resulting in small insertions-deletions (INDELs) and can be used for targeted gene disruption (32) (Figure 1).

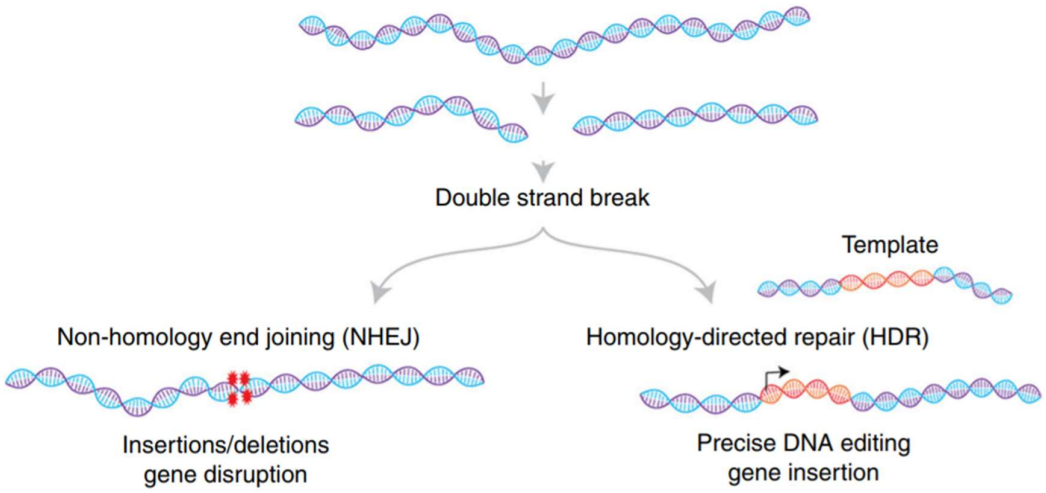


Figure 1. DNA DSB repair and possible outcomes. Chromosomal DNA DSB results in either NHEJ leading to indel mutations and gene disruption or HDR leading to precise gene editing. Taken from Adli, M. (33).

Synthesized short interfering RNAs (siRNAs) were used for targeted gene silencing, but validation of several thousand siRNAs used to target various genes demonstrated that many targets were difficult to knock down as the approach seriously lacked efficiency (34). Homing nucleases or meganucleases, which are rare-cutting endonucleases capable of recognizing long stretches of DNA, have also been used for targeted genome editing (35). For this approach, meganucleases for targeted cleavage have to be selected from libraries of variants (36). First sophisticated genome editing tools were programmable nucleases inducing targeted DSBs based on Zinc finger (ZF) proteins and Transcription Activator-like Effectors (TALEs), which both possess the required ability to target a specific DNA–binding site. These programmable nucleases both have the FokI nuclease domain C-terminally fused to differing DNA binding domains.

First engineered programmable nucleases were zinc-finger nucleases (ZNFs) with ZF protein domains, which were observed in transcription factor TFIIA form *Xenopus laevis* as DNA-binding motifs (37). ZF structural motifs consist of around 30 amino acids that fold into a specific structure where the Zinc ion stabilizes conserved Cys_2His_2 residues (2 Cys residues in β-sheets, and 2 His residues in α-helix). Each ZF interacts with a specific triplet it recognizes in the DNA sequence when its α-helix is inserted into the major groove of the double-helix (38). Since ZFs bind independently they can be linked in a peptide designed to bind a specific target site. ZFNs are fusions of zinc finger protein domains with C-terminal domain of FokI endonuclease from *Flavobacterium okeanokoites* which has a nonspecific DNA-cleavage activity (39). FokI nuclease dimerizes and dimers bind opposite DNA chains so two different ZFNs binding in close proximity (left and right ZFN) are required for targeted cleavage of one specific site (40) (Figure 2). Each of the monomeric ZFNs contains ZF modules recognizing specific triplets and binds to one of the two target half-sites separated by a 5–6 bp spacer sequence. Multiple monomeric ZFNs with different sets of ZFs can be designed for a single half-site (6). Individual pre-selected ZF domains have to be assembled when constructing ZFNs. Both natural ZFs with known triplet specificities and synthetic fingers targeting ANN, TNN, CNN and GNN are used (7,41). ZFNs usually have between 3 and 6 ZF motifs connected in a manner that enables them to bind target sequences from 9 to 18 base pairs in length (Figure 2). The success rate of modular assembly of ZF domains has been reported to be low and that remains a huge difficulty in construction of molecular tools

based on ZF proteins (42). Validation of 315 ZFN pairs targeting 33 sites in the *CCR5* gene showed that 44% ZFNs successfully cleaved target DNA *in vitro*, but only around 7% were efficient in cell-based reporter system. Higher success rates were observed for ZFN pairs in which each monomer contained four ZFs in comparison with pairs of ZFN monomers containing 3 ZFs each (6). ZFNs have been successfully used for silencing of the long non-coding RNA (lncRNA) MALAT1 in a manner that was more specific and 300x more efficient than before tried RNA interference approaches (43). ZFN off-target binding is a serious challenge and increasing the number of ZF motifs in modular assembly results in improved binding specificity, but makes the construction even more laborious.

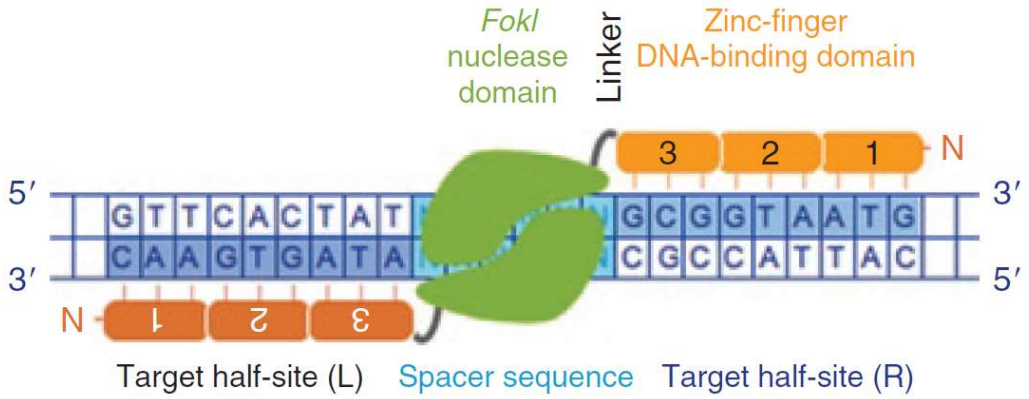


Figure 2. ZFN binding and DNA cleavage. A pair of ZFN monomers, each containing 3 ZFs (1-2-3) linked to the cleavage domain of FokI nuclease via short linker. Each finger interacts with a specific triplet, so each ZFN binds 9 bp within the target site. When both ZFNs bind, FokI can dimerize into active nuclease and cleave DNA at indicated spacer sequence, separating two target half-sites (L) and (R). Taken from Cathomen and Joung (44).

Significant step forward from ZFNs were Transcription activator-like effector nucleases (TALENs), which are fusions of TAL effector DNA-binding domains and FokI nuclease DNA cleavage domains (45). TALEs were identified from pathogenic bacteria *Xanthomonas*, which secrete these proteins to alter transcription of certain targeted genes in host plant cells to promote infection (46). Most investigated and commonly used TALE in genome editing is AvrBs3 from *Xanthomonas campestris* pv. Vesicatoria (47). TAL effector DNA-binding domain consists of repeats of conserved 33-34 amino acid sequence with variable 12th and 13th amino acids called Repeat Variable Diresidue (RVD). Each TALE

repeat recognizes a single nucleotide, its specificity determined by RVD (Figure 3). Secondary structure of TAL effector is predicted as two α-helixes flanking a loop domain in each repeat, RVD being inside the loop domain and accessing the DNA helix (46). TALE repeats are modularly assembled, like ZF domains, to target a specific DNA sequence (9).

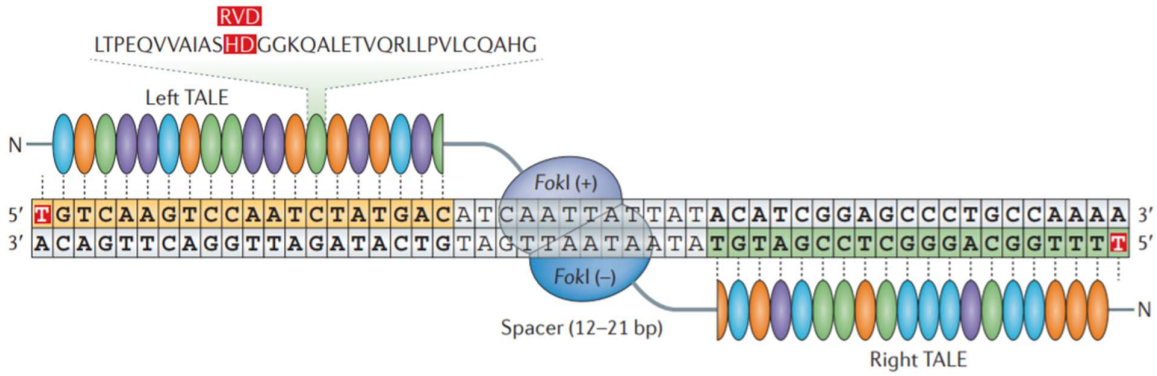


Figure 3. A pair of TALENs binding to the target site to enable FokI dimerization and DNA cleavage. Each TALEN is a fusion of TALEs at N-terminus (different coloured beads) and the FokI nuclease domain at C-terminus (blue shapes). Every TALE repeat consists of 33-35 amino acids, its specificity for base recognition determined by RVD at positions 12 and 13 (highlighted in red boxes). Total length of target sequence is typically 30-40 bp and for successful binding it must be preceded by T on each side (Ts in red boxes). When both TALENs bind, FokI can dimerize into active nuclease and cleave DNA inside the indicated spacer sequence. Taken from Kim and Kim (29).

Two common types of TALEs were tested in comparison with ZFN controls in a yeast-based *LacZ* reporter system and showed similar levels of targeted cleavage (45). Although TALENs and ZFNs might have comparable activity, TALENs are a better tool because of much looser targeting requirements since each TALE repeat recognizes a single base instead of a triplet (48). Because of this one-to-one correspondence between RVD modules and the four bases, TALENs can be designed to target almost any given DNA sequence which gives them a great advantage over ZFNs as many DNA sequences may lack targetable sites for ZFs (29). The only targeting limitation seems to be that target sites must be preceded by a thymine for TAL effector activity (49). Modular assembly of TALENs is time-consuming as they often consist of up to 20 RVDs, and even the most sophisticated methods

for Golden Gate assembly of TALENs developed require several cloning and selection steps in which fragments with specific compatible overhangs must be assembled for multiple RVDs in precise order (50).

2.2 CRISPR/Cas9 system

New systems recently emerged which greatly simplify the development of programmable nucleases, and these are the Clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems (10,51). The CRISPR/Cas systems play an important role in adaptive immunity of bacteria and archaea through guided cleavage of foreign nucleic acids. CRISPR loci consist of short palindromic repeats of CRISPR repetitive sequences and variable spacer sequences, next to c*as* genes. CRISPR spacer sequences are homologous to foreign DNA elements which CRISPR/Cas system recognizes and integrates in this form via *cas* genes, so the system works as a surveillance mechanism which recognizes foreign DNA, targets these homologous sequences with guide RNAs transcribed from the integrated spacer sequences and cleaves them through endonuclease activity (52–54). Discovered CRISPR/Cas systems can be roughly divided in two main classes and six types, differing in mechanisms of guide RNA processing, mature guide RNA length and Cas protein interactions. The class I CRISPR system is divided into types I, III and IV, which consist of large complexes of several coordinated proteins. The class II CRISPR system is divided into types II, V and VI, and these systems cleave foreign DNA with one RNA-guided nuclease (55,56). CRISPR-associated protein 9 (Cas9) is a part of class II CRISPR/Cas systems which works as an endonuclease guided by a short RNA molecule targeting a specific 20 bp sequence where it introduces a DSB. The Cas9 also requires a conserved sequence motif directly downstream of the targeted sequence, called protospacer-adjacent motif (PAM). CRISPR spacer sequences are transcribed into precursor CRISPR RNA (pre-crRNA) which is then cleaved into short CRISPR RNAs (crRNAs) targeting complementary protospacer sequences in foreign targets. For successful guidance of CRISPR/Cas systems, mature crRNA has to pair with a trans-activating crRNA (tracrRNA) to form a protein-RNA complex needed for targeted cleavage. In class II systems, pre-crRNA has to pair with tracrRNA in the presence of the Cas9 for its processing by endoribonuclease RNase III to occur (57,58)

(Figure 4), and it has been shown that fusing crRNA and tracrRNA into a single guide RNA molecule (sgRNA) successfully guides Cas9 as well (10).

The Cas9 protein from *Streptococcus pyogenes* SF370 type IIA CRISPR/Cas system (SpCas9) is the most commonly used in engineered CRISPR/Cas9 systems, as it has been studied in detail (57). This CRISPR/Cas9 locus consists of four *cas* genes, six spacer sequences flanked by repeats and a gene encoding a tracrRNA (Figure 4). SpCas9 recognizes a 5'-NGG-3' PAM sequence, which is quite frequent within the human genome (occurs every 8 bp). However, since both sense and antisense strands can be targeted for inducing a DSB, a possible PAM recognition sequence can be found once in every 4 bp, what makes the system ideal for genome editing in human cells. It is incredibly convenient to construct a programmable nuclease based on CRISPR/Cas9 system since one must simply design a sgRNA complementary to sequence of interest and deliver it together with Cas9, while for ZFNs and TALENs a new set of proteins must be engineered for each new target. Therefore, engineered CRISPR/Cas9 systems finally provide an affordable and easy approach to genome editing.

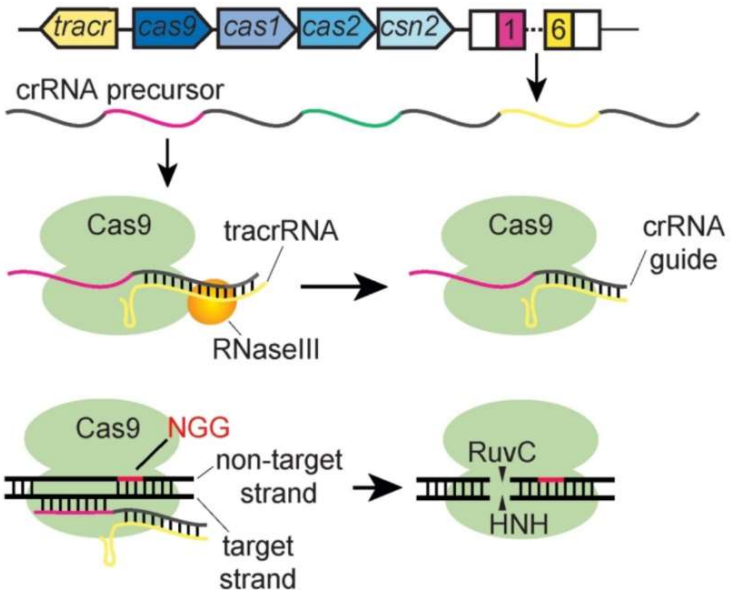


Figure 4. *Streptococcus pyogenes* SF370 CRISPR/Cas locus and CRISPR/Cas9 DNA cleavage. *Streptococcus pyogenes* SF370 CRISPR/Cas locus contains 4 *cas* genes (blue; *cas9, cas1, cas2, csn2*), tracrRNA gene (yellow; *tracr*), repeats (white boxes) and spacer sequences (coloured boxes 1-6). Spacer sequences are transcribed into pre-crRNA which pairs with

tracrRNA bound by Cas9. RNase III is recruited for cleavage of pre-crRNA forming Cas9-tracrRNA-crRNA complexes. These complexes scan dsDNA molecules for the presence of a PAM sequence (NGG). Following PAM binding, crRNA pairs with its homologous sequence and the Cas9 unwinds the DNA duplex and cuts the recognized foreign DNA. HNH nuclease domain cleaves the complementary strand (target strand) and RuvC nuclease domain cleaves the noncomplementary strand (non-target strand). Taken from Marraffini, L. (58).

First successful genome editing of human cells with engineered CRISPR/Cas9 system was done in HEK293 cells and the target was the *EMX1* gene (59). *EMX1* was disrupted by indel mutations resulting from NHEJ after Cas9 targeted cleavage, but Cas9 was also successfully converted into a nickase which facilitates HDR and specific point mutations were introduced to *EMX1* coding sequence resulting from HDR with a donor template with a specific single nucleotide mutation. Human codon-optimized Cas9 protein was developed and cloned into a mammalian expression system with a C-terminal SV40 nuclear localization signal (NLS). This system was successfully used for targeted cleavage in cell-based reporter system in HEK293T cells and for targeting of endogenous loci in K562 cells and PGP1 human induced pluripotent stem cells (iPSCs) (60). Purified Cas9 protein was successfully delivered to human leukaemia K562 cell line, embryonic stem cells (ESCs) and fibroblasts together with sgRNA in a RNP complex and efficiently induced targeted mutations. Primary cells are usually difficult to transfect, so direct delivery of RNP complexes might be a way to overcome the problem as it seems less stressful to sensitive cells. Such approach simplifies the construction and delivery of genome editing tools because simply replacing the sgRNA in the RNP complex makes the system ready for new targeting (61).

All three described types of programmable nucleases have been proved successful in targeted chromosomal DNA cleavage, but they differ in composition, targeting requirements, binding properties and limitations, off-target issues and other respects (29).

2.3 Epigenome editing

Disrupting genes by introducing indels followed by DSBs is precarious as it can potentially lead to off-target mutations, which is still a major concern in genome editing.

Epigenetic silencing can have the same repressive effect as gene disruption and CRISPR/dCas9 is a great platform for targeting of epigenetic effectors. Most of the complex, multifactorial diseases have a strong epigenetic component, so targeted epigenome editing is fundamental for elucidating their pathology. Targeted manipulation of epigenetic marks is also imperative for understanding the complexities of gene regulation in the context of responses to environmental stimuli, development and pluripotent stem cell self-renewal and differentiation (62). Direct functional studies of epigenetic modulation of gene expression have been limited by the lack of methods for targeted manipulation of epigenetic marks until development of platforms for DNA targeting. For example, effects of DNA methylation on gene expression could only be analysed after treatment of cells with non-specific methylation inhibitors, which is not a good approach as it results in global CpG demethylation of treated cells. ZFs, TALEs and CRISPR/Cas9 system, have been successfully repurposed in the last few years for recruitment of transcription factors and epigenome-modifying enzymes for targeted chromatin modifications. dCas9-based tools are becoming the most widely used platform due to its superior efficiency, specificity, versatility and ease of use (62). Epigenome editing might be necessary for downregulation of certain genes as it has been documented that, in some cases, successful gene targeting and knockout using TALENs was possible only after DNA methylation marks were removed from target sites using an inhibitor of DNA methyltransferase (50). Development and use of newly designed epigenome editing tools will provide information on how specific epigenetic marks are established and maintained through cell divisions. Fusions of catalytic domains of DNMTs and TET enzymes with ZFs, TALEs and dCas9 enabled targeted manipulation of methylation marks and have given new insights into the causal effects of specific CpG sites for gene expression regulation. Apart from DNMTs and TET CDs for targeted manipulation of DNA methylation, any activation or repression domain can be used in these systems for targeted modulation of transcription, as well as CDs of various epigenetic effectors for targeted chromatin modifications.

Recent studies include several successful examples of direct epigenome modification with molecular tools based on ZF proteins and TALEs. In one study, 223 yeast chromatin regulator proteins were fused to programmable ZFs and tested in a yeast reporter system, and also later recruited in combination with the VP16 transactivator to provide an even deeper insight into transcriptional mechanisms (63). Fusion of ZFs with DNMT3a catalytic domain has been

successfully used for targeted methylation of *SOX2* (TF implicated as oncogene in breast cancer) and *Maspin* (tumor suppressor usually methylated and silenced in metastatic breast cancer cells) gene promoters in human breast cancer lines SUM159 and MCF7 (64). Fusion of ZFs with DNMT3a CD successfully increased methylation levels of *EpCAM* gene promoter up to 26% in transiently transfected human ovarian cancer cells SKOV3 cells and reduced its expression levels, which are usually elevated in ovarian cancer cells (65). In another study, 7 different ZF proteins were fused with the transcriptional activator VP64 (a tetramer of the *herpes simplex* virus VP16 activator domain) creating artificial transcription factors (ATFs) used for targeting of *p16* locus in p16-active 293T cells and several of them successfully activated *p16* expression. Interestingly, stable genomic integration of one selected p16ATF into H1299 cells resulted in demethylation of *p16* CGI and trimethylation of histone H3K4, while trimethylation of H3K9 and H3K27 in the *p16* promoter were noted absent (66). This is a great example of how direct epigenome editing studies provide insights into interrelatedness and interdependence of epigenetic marks.

TALEs have a natural activation domain, but can be fused with more potent activation domains such as VP16 activation domain or VP64. Fusions of designed TALEs and VP64 activation domain significantly increased *VEGF-A* (*vascular endothelial growth factor A*) gene expression in HEK293 and primary BJ fibroblasts (67). Designer TALEs were synthesized and used for specific modulation of endogenous genes expression in 293FT cells and targeting of *SOX2* and *KLF4* gene promoters with these tools upregulated their expression by 5.5 ± 0.1 and 2.2 ± 0.1 folds, respectively (68). In a study by Maeder and collaborators, designer TALEs were fused with TET1 protein and TET1 CD and used for targeted demethylation of CpG sites within promoter region of *KLF4* gene in human K562 cells. Fusions with TET1 CD induced significantly higher demethylation changes than those with full length TET1 protein. One of the tested TALE-TET1-CD fusions reduced the methylation of CpGs located 10 and 16 bp from the 3' end of the construct binding site by 21% and 30%, respectively. Such designer TALE-TET1-CD fusions induced significantly higher levels of *RHOXF2/2B homeobox* and *HBB* (*human beta-globin*) genes when targeted to their promoter regions (69).

Streptococcus pyogenes CRISPR/Cas9 system has been repurposed for epigenome editing by inactivating Cas9 nuclease with induced mutations in active sites of both RuvC

(D10A) and HNH (H840A) cleavage domains generating the catalytically inactive dCas9. With abolished nuclease activity, CRISPR/dCas9 guided by sgRNA molecule can easily serve as a platform for delivery of various functional domains to gene promoters and enhancers, and any other regions of interest (Figure 5). Fusions of dCas9 and domains of different repressive chromatin modifiers such as KRAB (Krüppel associated box) effector domain and CS (Chromo Shadow) domain delivered with sgRNAs expressed from a murine RNA polymerase III U6 promoter enabled efficient transcriptional repression in human and yeast cells (12). In the same study, fusions of dCas9 and various activators such as VP64 or p65AD activated gene expression in HEK293 cells (12). In another study, a dCas9-VP64 expression plasmid was co-transfected with expression plasmids for 4 sgRNAs (individually and combined) targeting the *IL1RN* gene promoter in HEK293T cells and induced substantial gene activation. Combination of sgRNAs showed to be most efficient in gene activation, suggesting that targeting multiple sites in the promoter might provide most efficient gene activation (70). In our group, we successfully repurposed the CRISPR/Cas9 system for targeted DNA methylation by fusing dCas9 with a catalytic domain of human DNA methyltransferase 3A (DNMT3A) and achieved gene repression through targeted promoter CpG methylation (14). In a study by Stepper et al., dCas9 was fused with a Dnmt3a-Dnmt3l construct and this tool was successfully used for targeted DNA methylation at the *EpCAM*, *CXCR4* and *TFRC* gene promoters and subsequent gene repression. Methylation peaks were observed 25 bp upstream and 40 bp downstream of the PAM recognition sites (71). This is a nice example of how multimerization of protein complexes and such fusion design might provide an elegant way of achieving potent gene activation or repression effects.

TALE fusion with histone demethylase LSD1 successfully removed marks of active chromatin from targeted enhancers which resulted in downregulation of proximal genes (72). This is an interesting approach for unravelling the enhancer-target gene relationships. Fusion of dCas9 and the catalytic core of the human acetyltransferase p300, successfully acetylyzed histone H3 lysine 27 at targeted sites and activated gene expression from promoters and enhancers. Interestingly, for dCas9-p300 no synergistic activation effects were observed with multiple sgRNA targeting (13), which suggests different mechanisms of activation than in classic CRISPRa targeting. Development of epigenetic tools for targeting of different

chromatin modifiers to any cell model of interest, individually or in combination, will allow dissection of complex relationships between different chromatin marks.

sgRNA component of CRISPR/dCas9 system can also be repurposed for ncRNA delivery, in a manner that ncRNA sequence gets incorporated in sgRNA sequence without altering its DNA binding and dCas9 recruitment (62). Shechner and collaborators developed a platform named CRISPR-Display which incorporates large RNA segments in its sgRNA part and can be used for lncRNA function studies (73).

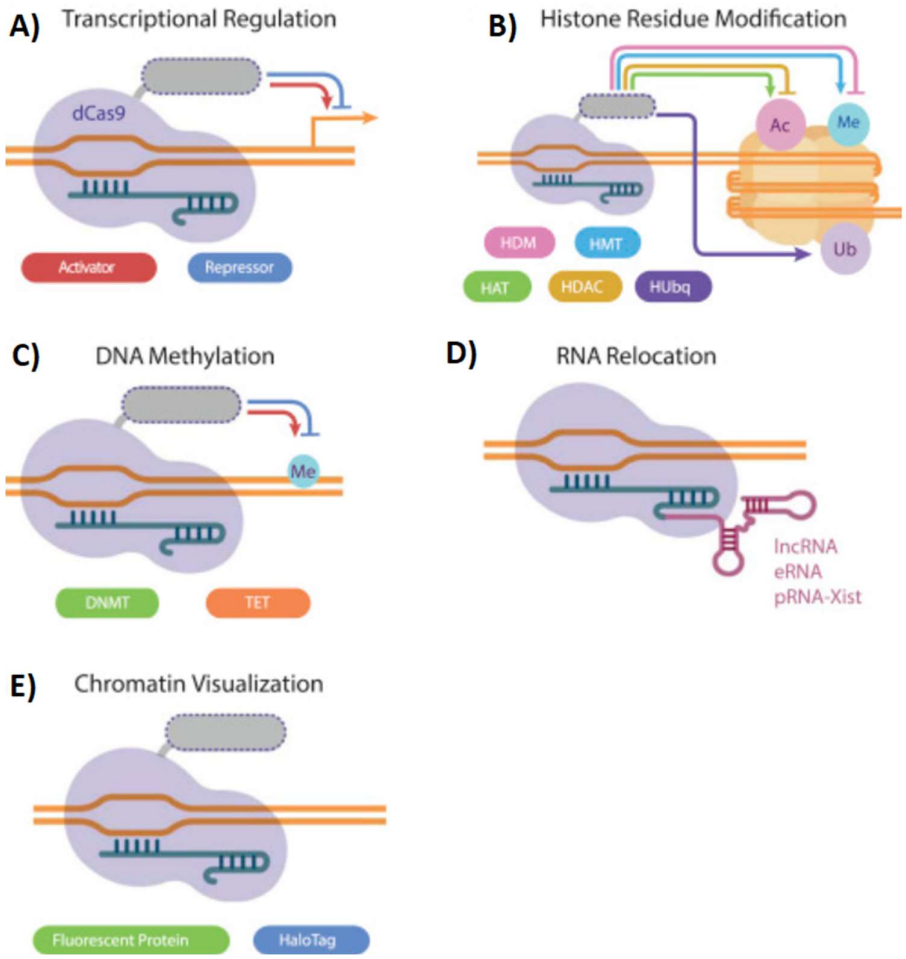


Figure 5. CRISPR/dCas9 systems for targeted epigenome editing. Chromatin modifiers and fluorescent proteins are modularly assembled with dCas9 for targeted manipulation of epigenetic marks or chromatin visualization. Grey dotted shapes indicate effector domains that can be replaced with modules of interest. **A)** various transcriptional activators or repressors (or their CDs) can be fused to dCas9 for transcriptional regulation; **B)** histone

modifiers (or their CDs) such as histone demethylase (HDM), histone methyltransferase (HMT), histone acetyltransferase (HAT); histone deacetylase (HDAC); histone ubiquitin ligase (HUbq) can be fused to dCas9 for inducing covalent histone modifications; **C)** enzymes responsible for DNA methylation and demethylation (or their catalytic domains) can be fused to dCas9 for targeted DNA methylation and demethylation; **D)** chromatin-interacting ncRNAs can be fused to dCas9 for RNA relocation; **E)** fluorescent proteins can be fused to dCas9 for targeted chromatin visualization. Taken from Pulecio et al. (62).

2.4 Epigenetic regulation of gene expression

Genetic information *i.e*. the nucleotide sequence in the DNA molecule, cannot be solely responsible for the multi-layered informational structure that makes a functional living organism. Epigenetics presents an additional layer of information, defining and ordering the underlying nucleotide sequence. Epigenetic mechanisms run gene regulation, determining when, where and in what manner will genes be expressed. The term epigenetics originates from the greek word *epigenesis* originally referring to differentiation processes in development and was created in 1942 by Conrad H. Waddington who described the *epigenotype* as a compilation of developmental processes between genotype and phenotype (74). Epigenetics is usually defined as mitotically and meiotically heritable changes in gene expression not encoded in the DNA sequence. These changes result from epigenetic mechanisms, encompassing DNA methylation, posttranslational histone modifications, chromatin remodelling, positional information, action of non-coding RNA molecules (ncRNAs) and histone variants. Epigenetic mechanisms are cooperative and interdependent. In animals, more than half of the genome is repressed in every cell of the organism, which suggests a sophisticated strategy for gene regulation and global mechanisms being in charge of maintaining gene expression patterns through cell divisions and development (4). DNA methylation refers to a covalent modification of cytosines in DNA molecule - an epigenetic modification important for the normal regulation of transcription, embryonic development, genomic imprinting, genome stability and chromatin structure. It was first proposed as a stably inherited epigenetic modification altering gene expression in 1975 (75,76) and it is the best known epigenetic mechanism since it has been easy to analyse even decades ago as there

were many known restriction endonucleases sensitive to cytosine methylation. DNA methylation in gene promoters is usually associated with gene repression, either through the interference of methyl group with transcription factor (TF) binding or through recruitment of methyl-CpG binding domain (MBD) proteins, which are part of repressive transcription complex consisting of histone deacetylases (HDACs) and histone methyltransferases (HTMs), responsible for repressive histone modifications, as well as of chromatin remodelers involved in formation of closed chromatin conformation (4). Covalent modifications of histones (acetylation, methylation, phosphorylation, sumoylation, ubiquitination, ADP-ribosylation and O-GlcNAcylation) are known to alter chromatin conformations and affect transcriptional states (77). Specific histone modifications are correlated with active and repressive chromatin states. For example, acetylation of histone H3 at position of lysine 27 (H3K27ac) is known as a mark of transcriptionally active chromatin, and tri-methylated histone H3 at position lysine 27 (H3K27me3) is associated with repressive chromatin states (78). Histone modifications are mostly readily reversible, while DNA methylation marks are considered more stable. Micro RNAs (miRNAs), siRNAs and lncRNAs can also modify chromatin and act as co-transcriptional silencing mechanisms (79). siRNAs guide Argonaute (AGO) and PIWI family of proteins to complementary RNA scaffolds and help recruit endogenous repressive mechanisms which induce DNA methylation and/or repressive histone modifications (79,80). The ENCyclopedia Of DNA Elements (ENCODE) project aims to identify and catalogue functional elements of the genome by using a variety of high-throughput methods (81). So far, ENCODE project discovered many novel non-protein-coding transcripts as well as previously unrecognised TSSs, and showed that histone modification patterns are predictive of presence and activity of TSSs (82). These, and other findings provided invaluable insights into the complex and dynamic epigenome.

2.5 DNA methylation and demethylation

DNA methylation is a process in which a methyl group ($-CH_3$) is added to the fifth carbon atom of cytosine generating 5-methylcytosine (5mC). In plants, cytosines can be methylated in CG, CHG or CHH contexts (H being A, T or C) (83). In mammals, more than 98% of DNA methylation occurs in CpG dinucleotides in somatic cells. Interestingly, in embryonic stem

cells (ESCs) nearly 25% of all methylation occurs in non-CpG sites (84). CpG dinucleotides are not equally distributed along the genome but are clustered into CpG islands (CGIs) which are on average 1000 bp long and usually unmethylated. Mammalian genomes are CpG-deficient because of the mutagenicity of 5mC itself (85), and most CGIs are kept hypomethylated, through not sufficiently understood mechanisms. High-throughput DNA sequencing of mouse and human genomes revealed similar CGI numbers per haploid genome (86). In both genomes around 50% of detected CGIs associated with annotated transcriptional initiation sites, the remaining 50% distributed equally between genes and within gene bodies. Those intergenic and intragenic CGIs are referred to as orphan CGIs due to their "mysterious" functional significance (86). Most CGIs are probably sites of transcriptional control, including orphan CGIs quite remote from annotated promoters (87). CpG methylation within gene promoters is usually correlated with gene transcription. 5mCs can directly interfere with transcription factor binding or they can be specifically bound by MBD proteins following recruitment of the machinery responsible for other epigenetic modifications, usually histone modifications, which create a transcriptionally repressive chromatin state (20,87). Some of the co-repressors recruited to methylated CpG sites are histone deacetylases, histone methyltransferases, as well as chromatin remodelling complexes (88).

DNA methylation is catalysed by specific enzymes called DNA-methyltransferases (DNMTs). In mammals, there are 3 main DNMTs: DNMT1, DNMT3A and DNMT3B. DNMTs catalyse the transfer of methyl group from the main donor of 5-mC - molecule S-adenosylmethionine (SAM) - to the 5th carbon atom of cytosine (Figure 6). DNMT3A and DNMT3B are essential for genome-wide *de novo* DNA methylation during development, while DNMT1 acts on hemi-methylated DNA and maintains the methylation pattern through DNA replication (90–92). DNMT3A and DNMT3B show equal affinity for hemi-methylated and non-methylated DNA (93). In living cells there is also DNMT3L, which is enzymatically inactive, however essential for establishment of methylation as it activates DNMT3A and DNMT3B by binding to their catalytic domains (94). DNMT3L recognizes unmethylated histone H3 lysine 4 and recruits and activates DNMT3 for *de novo* DNA methylation. DNMT2 is also a member of this protein family and is capable of cytosine methylation, but seems to act on tRNA rather than on DNA (95). Some insulator/boundary proteins (96,97)

and the RNAi repression system, are probably also involved in establishing and maintaining DNA methylation patterns, as well as MBD proteins and other chromatin-remodelling factors.

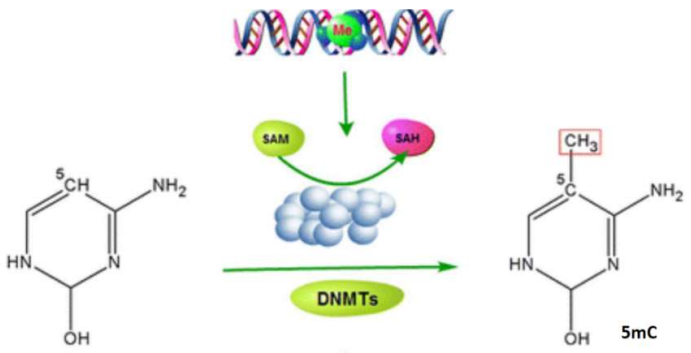


Figure 6. **DNA methylation by DNMTs.** DNMTs methylate the fifth carbon atom of cytosine generating 5mC. S-adenosylmethionine (SAM) is a main donor of methyl group which is then converted to S-adenosylhomocysteine (SAH). Adapted from Li et al. (89).

Establishment and maintenance of genomic methylation pattern plays an important role in organism development and normal cell function, with aberrant methylation status leading to development of various multifactorial diseases, including inflammatory and autoimmune diseases, diabetes and cancer (15,98). In mammals, DNA methylation pattern is established during embryogenesis and is maintained through DNA replications. DNA methylation and demethylation are part of the epigenetic reprogramming which happens in two waves, during pre-implantation development and during gametogenesis (99). Differential methylation of 1% of the genome in male and female gametes in mammals, called genomic imprinting, creates differing expression of so called imprinted alleles during development and results in functional differences between paternal and maternal genomes. DNA methylation pattern inherited from gametes is erased in pre-implantation development through genome-wide demethylation events. *De novo* methylation happens during the implantation of the embryo and the new pattern is established and maintained through future cell divisions (100).

In female mammals, one of the two X chromosomes is silenced during early embryogenesis during the X-chromosome inactivation process in order to assure dosage compensation of X-linked genes. Process is controlled by *Xic*, the X-chromosome-inactivation centre, and it includes coating the X chromosome by Xist RNA, DNA

methylation and histone modification (76,101). Another important function of DNA methylation is the silencing of transposons, which is crucial for genomic stability (102). Aberrations in the DNA methylation mechanisms have an important role in human disease. DNA methylation patterns are globally disrupted in cancer, with genome-wide hypomethylation and gene-specific hypermethylation events (especially in promoters of tumor-suppressor genes) occurring simultaneously in the same cell (103). Loss of normal imprinting leads to several human inherited genetic diseases (104,105) and mutations in MBD proteins can also alter methylation patterns and cause disease (106).

The cytosine residues methylated by DNMTs after DNA replication can be demethylated by passive "dilution" during subsequent replication or by active modification of the 5mC base. Passive DNA demethylation results from the errors which DNMT1 makes through maintenance of the methylation pattern during replication, which leads to the loss of some methyl groups (107). Active DNA demethylation requires enzymatic activity. There are several proposed mechanisms responsible for active DNA demethylation, involving various deaminases, DNA glycosylases which excise bases, enzymes involved in DNA repair and DNA-methyltransferases, but the pathways are not well understood and are actively investigated. Genome-wide demethylation was observed at specific moments of early development and gene-specific demethylation was observed in cells responding to various signals (99,108). First genome-wide demethylation occurs in the male pronucleus and it is rapid and independent of DNA replication, supporting the hypothesis that it happens actively. During this process, some genomic regions such as paternally imprinted genes and retrotransposons are protected from demethylation (99). Active DNA demethylation was first observed in extracts from murine erythroleukemia cells nuclei in 1982. Also, it seems that DNA demethylation can occur in cell-free environment with no DNA synthesis (109). Many mechanisms have been proposed as responsible for active DNA demethylation and it seems as it happens through cooperation of many different enzymes. MBD2 protein was proposed to be a DNA demethylase responsible for rapid DNA demethylation (110,111). Study by Metivier et al. proposed that DNMT3A and DNMT3B cyclically methylate Cs into 5mCs and demethylate 5mC through deamination (112). Discovery of a TET enzyme family has shifted the research focus to oxidation-mediated demethylation, and it seems that oxidation and DNA repair may be the key players in active DNA demethylation (21,113,114).

2.6 TET enzymes

The mammalian TET protein family contains TET1, TET2 and TET3. These are large multidomain enzymes with highly conserved C-terminal catalytic domains (115) (Figure 7). TET1, TET2 and TET3 were discovered as DNA demethylases through computational search for homologs of trypanosome enzymes JBP1 and JBP2 (21). JBP1 and JBP2 are members of 2-oxoglutarate- (2OG- or α-KG)- and Fe (II)-dependent oxygenase superfamily of enzymes found in trypanosomes which catalyse hydroxylation of the methyl group of T leading to a specific base J (β-D -glucosyl hydroxymethyluracil) formation. First, a specific T is converted into hydroxymethyldeoxyuridine (HOMedU), and then HOMedU is glycosylated to J (116), JBP1 and JBP2 are predicted to catalyse the first step (21). J is, like 5mC, associated with repressive chromatin states (116). Human TET1 (ten-eleven translocation 1) protein is capable to covert 5mC to 5hmC, and its overexpression in HEK293 cells results in reduced 5mC levels (21). TET1 also further oxidizes 5hmC to 5-formylcytosines (5fmC) and 5-carboxylcytosines (5caC), as well as TET2 and TET3 enzymes (Figure 8). Before the discovery of its role in DNA demethylation, TET1 was detected in fusions with MLL in acute myeloid leukaemia (117).

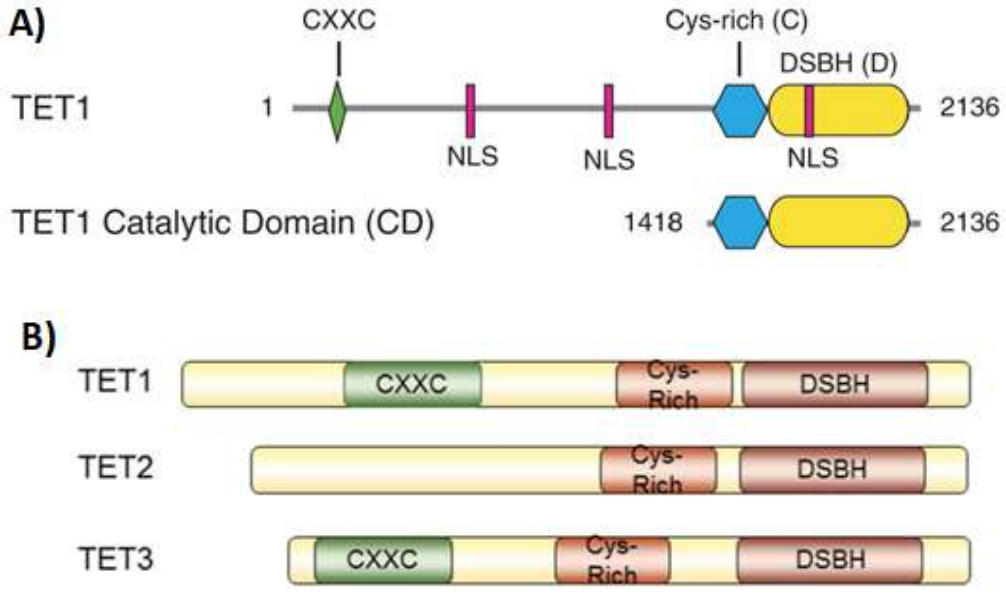


Figure 7. Schematic representation of TET enzymes. A) Human TET1 consists of CXXC-type zinc-binding domain which binds CGIs, the C-terminal catalytic domain (CD) containing cysteine-rich region (Cys-rich, C) and the double stranded β-helix domain (DSBH, D) which

have 2-oxoglutarate(2OG)- and iron(III)-dependent methylcytosine dioxygenase activity, and three nuclear localization sequences (NLSs). Taken from Tahiliani et al. (21). **B)** TET1, TET2 and TET3 have almost identical CDs, differing only in small spacer region length. TET2 differs from TET1 and TET3 because it does not have a N-terminal CXXC DNA-binding domain. Taken from Kao et al. (25).

TET-mediated demethylation of 5mC is an enzymatic reaction which requires Fe(II) and 2OG as substrates. TET1 actively removes 5mC via oxidation leaving 5hmC which can be turned into unmethylated cytosine by subsequent oxidation and repair. 5hmCs are further oxidized by TET proteins to 5fmC and 5caC, which are turned in to unmethylated cytosines after thymine-DNA-glycosylase (TDG) recognizes G:5fC and G:5caC, excises them in the base excision repair (BER) process and leaves an abasic site (118) (Figure 8). 5hmC, 5fC and 5caC could facilitate replication-dependent DNA demethylation. UHRF1 recognizes hemimethylated CpG sites after replication and interacts with DNMT1 which restores symmetrical CpG methylation. Modified cytosines - 5hmC, 5fC and 5caC - could interfere and block UHRF1 binding and therefore lead to loss of CpG methylation through replication (119).

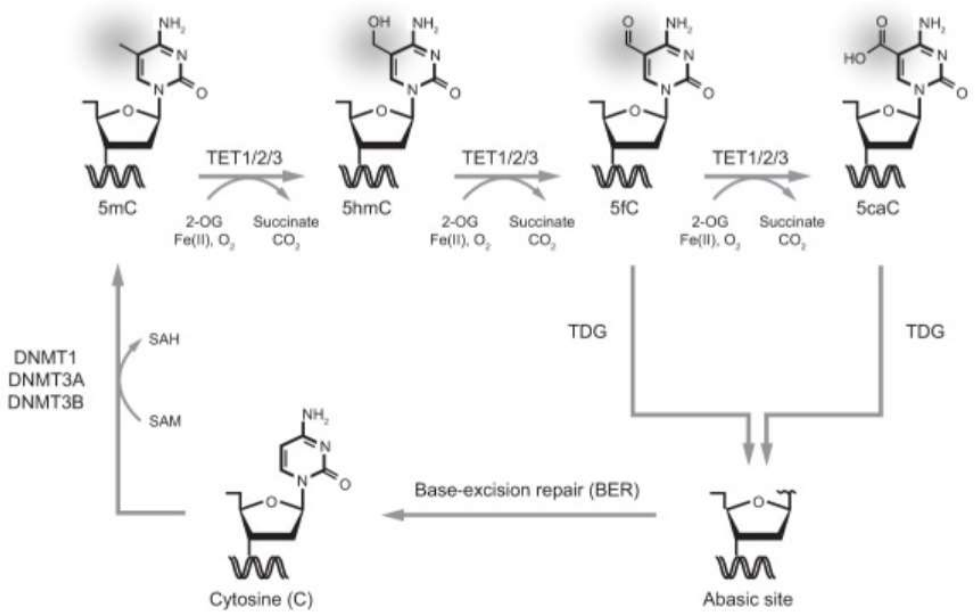


Figure 8. Active DNA demethylation by TET enzymes. DNMTs methylate the fifth carbon atom of cytosine generating 5mC, using SAM as methyl group donor which is converted to

SAH. 5mC can be sequentially oxidised by TET proteins to generate 5hmC, 5fC and 5caC, which can be lost through passive demethylation through cell division or 5fC and 5caC can be excised by TDG, which will then be repaired by the base excision repair mechanisms and converted to unmethylated C. Taken from Rasmussen and Helin (120).

DNA demethylation is more complex to study than DNA methylation because conventional bisulfite sequencing methods are not precise in base definition when it comes to C, 5fC, 5caC, 5hmC and 5mC. Bisulfite conversion is treatment of DNA with sodium bisulfite, which causes cytosine deamination to uracil, while 5mC does not react with bisulfite and remains resistant to deamination during treatment. Cytosine, 5fC and 5caC are all deaminated to uracil during sodium bisulfite treatment and are later converted to, and sequenced as T, while both 5hmC and 5mC are sequenced as C (25). Specific product of 5hmC reacting with bisulfite is called cytosine 5-methylenesulfonate (CMS), and although there is no difference between sequencing of 5mC and 5hmC after conversion (both are read as C), it has been found that CMS rich DNA regions were less efficiently amplified than regions containing only 5mC or U after conversion. CMSs may stall DNA polymerases in PCR, therefore 5hmC rich DNA regions might be underrepresented in quantitative methylation analyses (121). Without fine distinction between modified bases which are intermediates and products of DNA demethylation it is impossible to precisely study their possible regulatory roles. Development of oxidative bisulfite (oxBS) conversion methods provides the best available approach for 5hmC detection (122). These methods aim to selectively oxidize 5hmC to 5fC, which is then bisulfite converted to uracil in later sodium bisulfite treatment. Exact role of 5hmC remains unclear, although several studies imply on specific regulatory functions (*i.e.* the 6th base in the DNA; 5mC being the 5th base), rather than being just an intermediate residue in DNA demethylation pathways. 5hmC was discovered as a modification in genomes of bacteriophages, making the viral DNA unsusceptible to degradation by host defence mechanisms (123), therefore it seems likely that this base has an important function in complex genomes as well. Stable isotope labelling of modified C bases in mammalian DNA paired with liquid-chromatography mass spectrometry showed that most of 5hmC might be an independent stable modification, as it takes 30 hours following DNA synthesis for 5hmC to form, while 5mC form during replication (124). 5hmC is thought to be tissue specific - it was found to be around 40% as abundant as 5mC in DNA of Purkinje

neurons (125). TET1 was found to be critical for neuronal activity and memory functions via regulation of DNA methylation levels at promoter regions of crucial genes involved in the transcriptional regulation of cognitive processes (126).

TET enzymes and 5hmCs have important roles in differentiation during development, and although their presence and function in placental epigenome were thoroughly investigated by several studies, there are still many unanswered questions. One study identified around 21 000 loci with higher levels of 5hmC in the placental genome. 5hmC was found to be absent in CGIs and depleted in CpG shores, but enrichment of 5hmC was noted in CpG open sea and shelf regions (127). A genome-wide map of 5hmC in human ESCs was created by hydroxymethyl-DNA immunoprecipitation followed by deep sequencing (hmeDIP-seq), and 5hmC enrichment was detected in enhancers and gene bodies (128). 5hmC was found in undifferentiated ES cells, but not in dendritic cells or T cells, and the same differentiated ESCs showed a 40% decrease in 5hmC levels. Same study showed that TET1-depleted ESCs experienced a significant decline in 5hmC levels TET1 mRNA levels and that TET1 mRNA levels drop drastically in ESCs when differentiation is induced (21). *Tet1* transcripts have been detected in mouse ESCs at similar levels as pluripotency factor *Oct4*, *Tet2* being 5-fold less abundant and *Tet3* at very low levels, suggesting that TET1 and TET2 produce majority of 5hmC and guide ESCs differentiation (129). TET1 was implied to be crucial in mouse ESCs self-renewal through maintaining the expression of *Nanog* pluripotency factor (130). *Tet1* mutant mice and mutant mouse ESCs were generated to investigate the role of *Tet1* in pluripotency and development. Such ESCs had reduced levels of 5hmC and certain global gene expression changes, but remained pluripotent although with slightly altered differentiation patterns. *Tet1* mutant mice were a bit smaller than controls, but were viable and fertile (131). Since all TET proteins have similar enzymatic activity, their roles can be considered redundant and loss of one TET member might show significant phenotype-level changes only in tissues unavailable to other TET enzymes (132). A *Tet1/Tet2/Tet3* knockout mouse ESCs were generated and showed depleted 5hmC levels and poor differentiation (133). 5hmC levels were found to be drastically changing in gene bodies during T-cell differentiation, higher levels of 5hmC correlating with higher gene expression and loss of 5hmC correlating with decrease in gene expression (134).

TET1, TET2, and TET3 also catalyse the subsequent conversion of 5hmC to 5fC and 5caC and these bases are present in 10–100-fold lower amounts than 5hmC (120), but it is possible that 5fC and 5caC might have specific regulatory roles as well. Proteins which specifically bind to 5fC have been discovered, suggesting its regulatory significance (135). Jin et al. suggest that in differentiated cells TET1 might work as a maintenance DNA demethylase, not purposely decreasing 5mC levels for transcriptional changes, but rather preventing *de novo* methylation spreading from the edges of methylated CGIs into surrounding regions. This would suggest that TET1 has a role in maintenance of normal methylation pathways post-development, and it sheds a new light on methylation deregulation in certain diseases and aging (136). Aberrant DNA demethylation has an important role in human diseases. Loss of 5hmC has been observed in cancers, as well as mutations in all three *TET* genes (137). *Tet2* was found to be a critical tumor suppressor in myeloid malignancies in mice (138), and *TET2* is known as one of the genes mutated in the early development of hematopoietic malignancies (120). Aberrant DNA demethylation of certain oncogenes can lead to their activation and carcinogenesis. For example, *synuclein gamma* (*SNCG*) gene is overexpressed in gastric cancers and advanced-stage breast and ovarian cancers, and its protein product stimulates proliferation of cancer cells and metastasis. The underlying mechanism of this aberrant expression is promoter CGI demethylation (139).

3. Materials and methods

3.1 Materials

Commercial kits:

Charge Switch PCR Clean-Up Kit (Invitrogen, Carlsbad, California, USA), Gel Elute Gel Extraction Kit (Sigma-Aldrich, Saint Louis, Missouri, USA), Quick Ligation Kit (NEB, Ipswich, Massachusetts, USA), QIAprep Spin Miniprep Kit (Qiagen, Venlo, Netherlands), QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, California, USA), Emerald Amp HS PCR Master Mix (Takara, Kusatsu, Shiga, Japan), DNeasy Blood & Tissue Kit (Qiagen), PyroMark PCR Kit (Qiagen), PyroMark Q24 Advanced CpG Reagents (Qiagen), EZ DNA Methylation-Gold Kit (Zymo Research Europe, Freiburg, Germany), RNA/DNA Purification Kit (Norgen Biotek, Thorold, Canada), RNeasy Mini Kit (Qiagen), TURBO DNA-free Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA), 2x TaqMan Gene Expression Master Mix (Applied biosystems, Foster City, California, USA), TaqMan Gene Expression Assays: Hs02379589_s1 (*MGAT3*), Hs01369240_m1 (*MGAT3*), Hs02800695_m1 (*HPRT1*) and Hs00609297_m1 (*HMBS*) (Applied Biosystems), Mini Quick Spin Columns (Roche, Basel, Switzerland)

Enzymes and buffers:

DNA polymerase Herculase II fusion DNA polymerase (Agilent Technologies), EcoRI (NEB), SphI (NEB), T4 DNA ligase (Takara); T4 DNA ligase Buffer (Takara), BamHI-HF (NEB), FseI (NEB), T4 Polynucleotide Kinase (NEB), BbsI (BpiI) (Thermo Scientific), Buffer G (NEB), RecBCD (NEB), PrimeScript Reverse Transcriptase (Takara)

Chemicals used in cell culture:

Trypsin – EDTA 0.25% solution (Sigma-Aldrich), OptiMEM (Thermo Fisher Scientific), Lipofectamine 3000 Transfection Reagent (Invitrogen), 1 × PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM K_2HPO_4, pH 7.4), puromycin (Thermo Fisher Scientific)

Other chemicals:

Ampicillin (Sigma-Aldrich), Luria-Bertani (LB) liquid media (10 g/L NaCl, 5 g/L tryptone,10 g/L yeast extract), MassRuler DNA Ladder Mix (Thermo Fisher Scientific), 100 bp DNA Ladder Ready to Load (Solis Biodyne, Tartu, Estonia), GelPilot DNA Loading Dye 5× (Qiagen), Sybr Green Gel Stain (Sigma-Aldrich), PyroMark Denaturation Buffer (Qiagen), PyroMark Wash Buffer (Qiagen), Streptavidin Sepharose High Performance Beads (GE Healthcare, Chicago, Illinois, USA), Random hexamer primers (Invitrogen), 10 mM dNTP Mixture (Sigma-Aldrich), Recombinant RNase Inhibitor (Takara), Potassium perruthenate $KRuO_4$ (Alfa Aesar, Haverhill, Massachusetts, USA)

Bacterial strain:

XL10GOLD chemically competent *E. coli* strain (Agilent Technologies)

Model cell line:

HEK293

3.2 Construction of dCas9-TET1 molecular tool for targeted CpG demethylation

pJFA344C7 (plasmid #49236, Addgene plasmid repository), a TET1-catalytic domain expression vector for TALE-TET1 fusion protein to induce targeted demethylation in mammalian cells was used as a source of TET1 catalytic domain of human *TET1* gene (tet methylcytosine dioxygenase 1) (69). TET1 catalytic domain was PCR amplified from pJFA344C7 using primers which contain restriction enzyme sites for BamHI and EcoRI (forward primer) and FseI and SphI (reverse primer), as well as additional overhangs (5 bp on 5' end of forward primer and 3 bp on 3' end of reverse primer) to ensure efficient digestion (Table 1). For this PCR amplification a high-fidelity DNA polymerase Herculase II fusion DNA polymerase (Agilent Technologies) was used, with following cycling conditions: initial denaturation at 95°C for 2 min; 30 cycles of denaturation at 95°C for 20 s, annealing at 60°C for 20 s, elongation at 72°C for 2 min; final elongation at 72°C for 3 min. Unique PCR product of 2188 bp was confirmed with agarose gel electrophoresis and Sybr Green (Sigma-Aldrich) staining and purified using Gel Elute Gel Extraction Kit (Sigma-Aldrich).

PCR product and plasmid pUC19 were cut with EcoRI and SphI (NEB), products were purified from double digestion reaction mixtures using Charge Switch PCR Clean-Up Kit (Invitrogen), ligated with Quick Ligation Kit (NEB) for 10 min at room temperature in 3:1 molar ratio of insert to vector, and transformed into XL10GOLD chemically competent *E. coli* strain (Agilent Technologies). Briefly, XL10GOLD were thawed on ice for 10 min, 5 µL of ligation mixtures were added and suspensions were gently mixed and then heat-shocked at 42°C for 35 s and briefly cooled on ice. Warm LB media was added to transformed XL10GOLD and they were incubated at 37°C for 30 min while shaking at 350 rpm. Bacterial suspensions were plated on warm LB plates containing 100 µg/mL ampicillin. Plates were incubated for 16 h at 37°C.

Colony PCR was performed with Emerald Amp HS PCR Master Mix (Takara) and primers TET1_BamHI_Fw and TET1_FseI_Rev under cycling conditions: initial denaturation at 98°C for 2 min; 30 cycles of denaturation at 98°C for 10 s, annealing at 60°C for 30 s, elongation at 72°C for 3 min; final elongation at 72°C for 2 min. Plasmids, named pUC19-TET1_CD, were isolated from several positive transformants cultured for 16 h in 4 ml of liquid LB-Amp (100

μg/mL) cultures, using QIAprep Spin Miniprep Kit (Qiagen) and verified by Sanger sequencing.

TET1 catalytic domain originally contains 2 BbsI restriction sites, which would interfere with subsequent cloning, so they were mutated using 2 pairs of mutagenic primers (Table 1) and QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent Technologies), in a way that TET1 protein sequence remains unchanged.

Catalytically inactive version of TET1 catalytic domain named DED1 variant was created by introducing two mutations, H1671Y, D1673A, into TET1 active site with one pair of mutagenic primers (Table 1) in the same reaction with 2 pairs of mutagenic primers designed to eliminate BbsI sites. These substitutions have been found to abolish TET1 demethylation activity (21), most likely by impairing iron binding to active site. Successful mutageneses of pUC19-TET1_CD and pUC19-DED1_CD were confirmed by Sanger sequencing of entire catalytic domains with 4 designed sequencing primers (Table 1).

Plasmid pSpCas9n(BB)-2A-Puro (plasmid #48141, Addgene plasmid repository) (11) with additional mutation in HNH domain of Cas9 was used as backbone for this modified CRISPR/Cas9 system. Mutated pSpCas9n(BB)-2A-Puro, with substitutions in active sites of both RuvC (D10A) and HNH (H840A) domains of Cas9, expresses Cas9 in completely catalytically inactive form, "dead" Cas9 (dCas9). Finally, a short Gly_4Ser linker with BamHI and FseI restriction sites was added to the C-terminus of the dCas9 gene as described in previous work from our lab (17), and such plasmid variant, named pdCas9-GS, was used as backbone for insertion of TET1(DED1) catalytic domain.

Both pdCas9-GS and pUC19-TET1(DED1) _CD were cut with BamHI-HF and FseI (NEB), digestion reaction mixtures were visualized by agarose gel electrophoresis with Sybr Green (Sigma-Aldrich) staining and fragments of correct size were purified using Gel Elute Gel Extraction Kit (Sigma-Aldrich). Vector and inserts were ligated with Quick Ligation Kit (NEB) for 10 min at room temperature in 3:1 molar ratio of insert to vector, and transformed into XL10GOLD chemically competent *E. coli* strain (Agilent Technologies). To select for positive clones, colony PCR was performed with Emerald Amp HS PCR Master Mix (Takara) and primers TET1_BamHI_Fw and TET1_FseI_Rev as described previously. Plasmids, named dCas9-GS-TET1_CD and dCas9-GS-DED1_CD, were isolated from positive

transformants using QIAprep Spin Miniprep Kit (Qiagen) and verified by Sanger sequencing. It was essential to check for any possible changes in open reading frame, as BamHI triplet acts as a start codon for expression of TET1/DED1 catalytic domain and GS-linker-seq primer was used for this verification (Table 1).

Finally, selection markers were cloned at the C-terminus of the dCas9-GS-TET1(DED1) fusion to allow for future successful manipulation of cells transfected with this molecular tool. Plasmid pSpCas9n(BB)-2A-GFP (plasmid #48140, Addgene plasmid repository) (11) was cut with EcoRI (Thermo Scientific) and used as source of T2A-EGFP fragment. Mutation H166R was introduced into puromycin resistance gene of pSpCas9n(BB)-2A-Puro using a pair of mutagenic primers (Table 1) and QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies), to achieve improved puromycin selection like the one in plasmid pSpCas9n(BB)-2A-Puro V2.0 (plasmid #62987, Addgene plasmid repository) (11). Successful mutagenesis was confirmed by Sanger sequencing of puromycin gene with BGH-R sequencing primer (Table 1). EcoRI fragment encoding this T2A-Puro_H166R was cut from mutated pSpCas9n(BB)-2A-Puro, visualized on agarose gel and gel purified as described previously. Vectors dCas9-GS-TET1_CD and dCas9-GS-DED1_CD were cut with EcoRI (Thermo Scientific), purified from digestion reaction mixtures using Charge Switch PCR Clean-Up Kit (Invitrogen), and dephosphorylated to prevent vector re-circularization using Fast AP (Thermo Scientific) according to manufacturer's instructions. EcoRI T2A-EGFP and T2A-Puro_H166R fragments were ligated with dephosphorylated EcoRI cut vectors dCas9-GS-TET1_CD and dCas9-GS-DED1_CD using Quick Ligation Kit (NEB) as described previously and transformed into XL10GOLD (Agilent Technologies). Colony PCR was performed with Emerald Amp HS PCR Master Mix (Takara) and primers TET1_seq3 and GFP-in-rev or PuroR-in-rev (Table 1) under cycling conditions: initial denaturation at 98°C for 2 min; 30 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 30 s, elongation at 72°C for 1 min; final elongation at 72°C for 1 min. Plasmids, named dCas9-TET1(DED1)-T2A-EGFP and dCas9-TET1(DED1)-T2A-PuroR, were isolated from positive transformants using QIAprep Spin Miniprep Kit (Qiagen) and verified by Sanger sequencing.

Maps of constructed tools dCas9-TET1(DED1)-T2A-PuroR and dCas9-TET1(DED1)-T2A-EGFP are shown in Figure 12. Maps of vectors used as source material and shuttle vectors are shown in Appendix 1. Sequences of all primers used in cloning (PCR

amplifications, mutageneses and sequencing reactions) are listed in Table 1. All primers were commercially synthesized and purified by Macrogen Oligo DNA synthesis service. All Sanger sequencing reactions were performed by Macrogen DNA sequencing service. All sequencing results were analysed with SnapGene (GSL Biotech, Illinois, USA).

Table 1. Primers used in pdCas9-TET1(DED1)-T2A-PuroR and pdCas9-TET1(DED1)-T2A-EGFP construction

Name	Sequence (5'→ 3')	Use
TET1_BamHI_Fw	ATCTAGAATTCGGATCCCTGCCCACCTGCAG	PCR amplification of TET1 CD from pJFA344C7, introducing restriction sites and overhangs
TET1_FseI_Rev	TCAGCATGCGGCCGGCCGACCCAATGGTTATAGGGCCCCGC	
TET1_no1stBbsI_S	CTTCTCCTGGTCCCCAAAGACTGCTTCAGCC	mutagenesis of 1st BbsI restriction site in TET1(DED1) CD
TET1_no1stBbsI_AS	GGCTGAAGCAGTCTTTGGGGACCAGGAGAAG	
TET1_no2ndBbsI_S	GATGCCTTCGGGAAGGCTCAGTGGTGCCAAT	mutagenesis of 2nd BbsI restriction site in TET1(DED1) CD
TET1_no2ndBbsI_AS	ATTGGCACCACTGAGCCTTCCCGAAGGCATC	
TET1toDED1_S	GACTTCTGTGCTCATCCCTACAGGGCCATTCACAACATGAATAA	mutagenesis of active site in TET1 CD, introducing H1671Y, D1673A
TET1toDED1_AS	TTATTCATGTTGTGAATGGCCCTGTAGGGATGAGCACAGAAGTC	
TET1_seq1	GTTTGGCTACACGATTAGCTCC	sequencing of entire TET1(DED1) CD
TET1_seq2	CTTATTCGCTGATGCCATCCG	
TET1_seq3	GCACCCCAACCGTAATCATC	
TET1_seq4	CATGTAAGGGAGAGCTTGGATC	
GS-linker-seq	GAAGAGGTACACCAGCACCAAAG	sequencing of NLS, Gly_4Ser linker and first codons of TET1(DED1) CD
PuroR_H166R_S	GAGACCTCCGCGCCCCGGAACCTCCCCTTCTAC	mutagenesis of puromycin resistance gene, introducing H166R

PuroR_H166R_AS	GTAGAAGGGGAGGTTCCGGGGCGCGG AGGTCTC	
BGH-R	TAGAAGGCACAGTCGAGG	sequencing of puromycin resistance gene
GFP-in-rev	TGGTGCAGATGAACTTCAGG	colony PCR; verification of successful cloning of T2A-EGFP or T2A-Puro_H166R in right orientation at the C-terminus of the dCas9-GS-TET1(DED1)
PuroR-in-rev	CGTGAGGAAGAGTTCTTGCAG	

3.3 Selection of loci for targeting with dCas9-TET1 and sgRNA design

Highly methylated promoter regions of targeted *MGAT3* and *LAMB1* genes in HEK293 cells suitable for targeted DNA demethylation were selected based on previous pyrosequencing results obtained in our lab. For the *MGAT3* gene two assays for bisulfite sequencing were used, MGAT3-A1 and MGAT3-A2, covering a total of 14 individual CpG sites (111 bp span). For the *LAMB1* gene assay LAMB1-A1 was used for analysis of 6 CpG sites (38 bp span). Assay sequences are listed in Table 2, assay maps are shown in Figures 10 and 11.

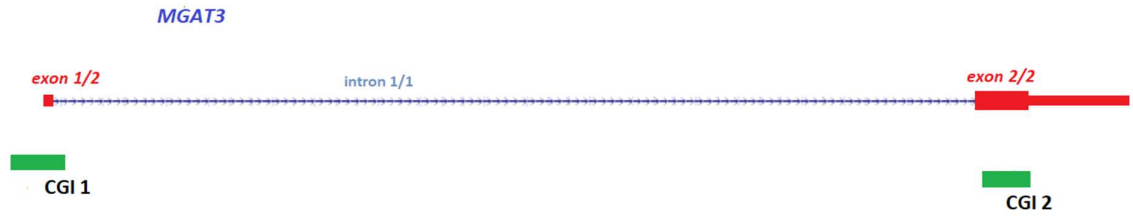
Guide sequences targeting the *MGAT3* and *LAMB1* loci were manually selected. Few parameters were crucial in guide selection: (i) sequences had to be highly specific, (ii) length of 20 bp was desired, (iii) sequences had to be followed by an immediate NGG PAM sequence on 3' end. Positions of selected sgRNAs are shown in Figures 9 and 10. Non-targeting (NT) control guide, non-homologous to any sequence in the human genome, was taken from the Human GeCKOv2 Library (140).

Table 2. Sequences of pyrosequencing assays used for methylation analysis of the *MGAT3* and *LAMB1* fragments.

Assay name	Assay sequence 5'→3' (analysed CpG sites are underlined)
MGAT3-A1	GCTGGGATATAGAATAGGTAG**CG**CATCCCTGCACCTT**CG**A**CG**ATGG**CG**G**CG**CAG AGATGTCTGCTG**CG**TACCCACAATGCCTTGTGCCT**CG**CAC**CGCG**GGAGGAAGTG GCTGCTCTGTAGGCCCCAGAACGGAACCACTTGAAAGGCGGGAAACACGTGGGG GACGCCTCTGAGCCCTGAGAGGAATGGC
MGAT3-A2	GCCTCTGAGCCCTGAGAGGAATGGCCTAGAGCAAGGCCA**CG**AGGAGCCA GGGCA**CG**ACA**CG**GTGGGCCCT**CG**GAGAAC**CG**CTGGTGGGCAAGTGGCA GGAGAGTAGGCTCAAGAGGGT
LAMB1-A1	GAAGTGGAGGGTCTACACATCCACCCTTTGTTGGGGGAGCTGCTCCCCGG GGCTGATCAGGGTGGG**CGCG**AG**CG**T**CG**GTGGGTTTCC**CG**GGAGGGAGG CTCT**CG**CTGCTGGACAGACCTGAT

A)

B)



C)

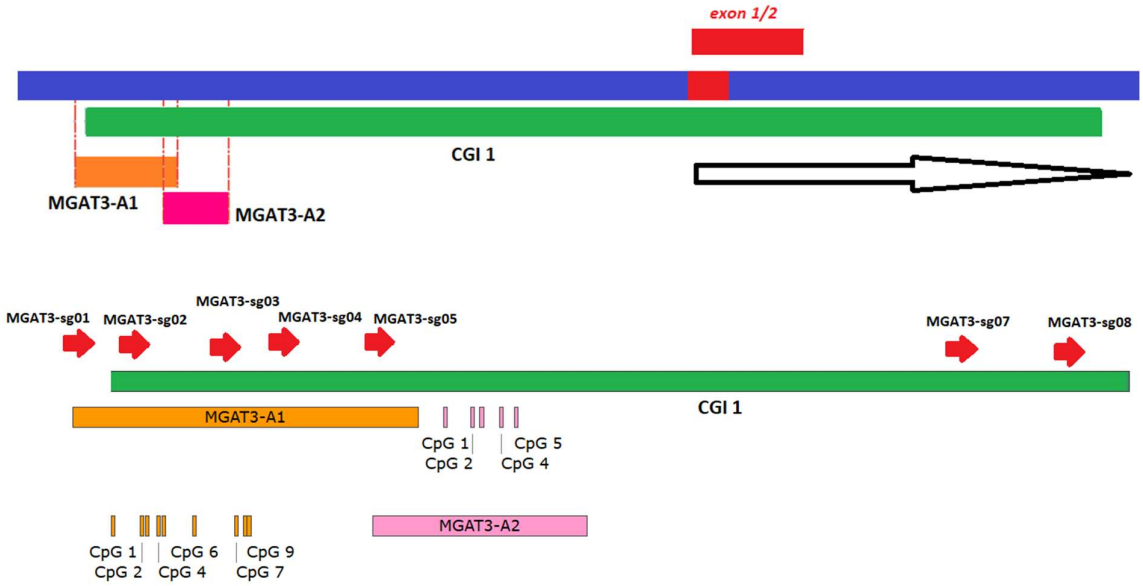
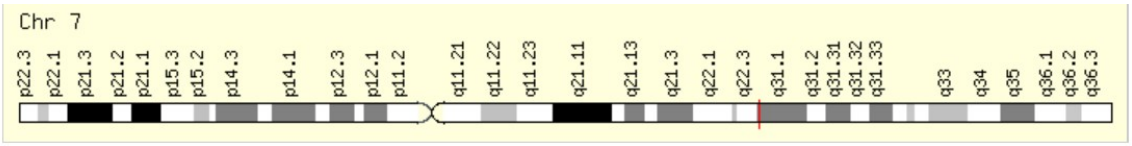


Figure 9. *MGAT3* gene location, scheme and sequence assayed for CpG methylation. **A)** *MGAT3* gene position on human chromosome 22 is marked with a red line on the chromosome scheme (q13.1). Representation taken from *www.genecards.org*., bands according to *Ensembl*. **B)** *MGAT3* gene scheme is shown with its 2 exons (red boxes) and intron (blue line) and 2 CGIs, as represented in GRCh38/hg38. **C)** Assays MGAT3-A1 and MGAT3-A2 are located immediately upstream and in the beginning of the first CGI (as indicated in the enlarged picture below, arrow indicating the transcript orientation). In

GRCh38.p12 primary assembly, this CGI stretches from chr22 39456201 – 39457987 (+). In additionally enlarged picture, assays MGAT3-A1 and MGAT3-A2 with 9 analysed CpG sites in MGAT3-A1 (orange lines, CpG 1-CpG 9) and 5 analysed CpG sites in MGAT3-A2 (pink lines, CpG 1-CpG 5) are shown with indicated positions of sgRNAs chosen for dCas9-TET1 targeting (MGAT3-sg01-MGAT3-sg08; red arrows indicating their position and orientation).

A)



B)

C)

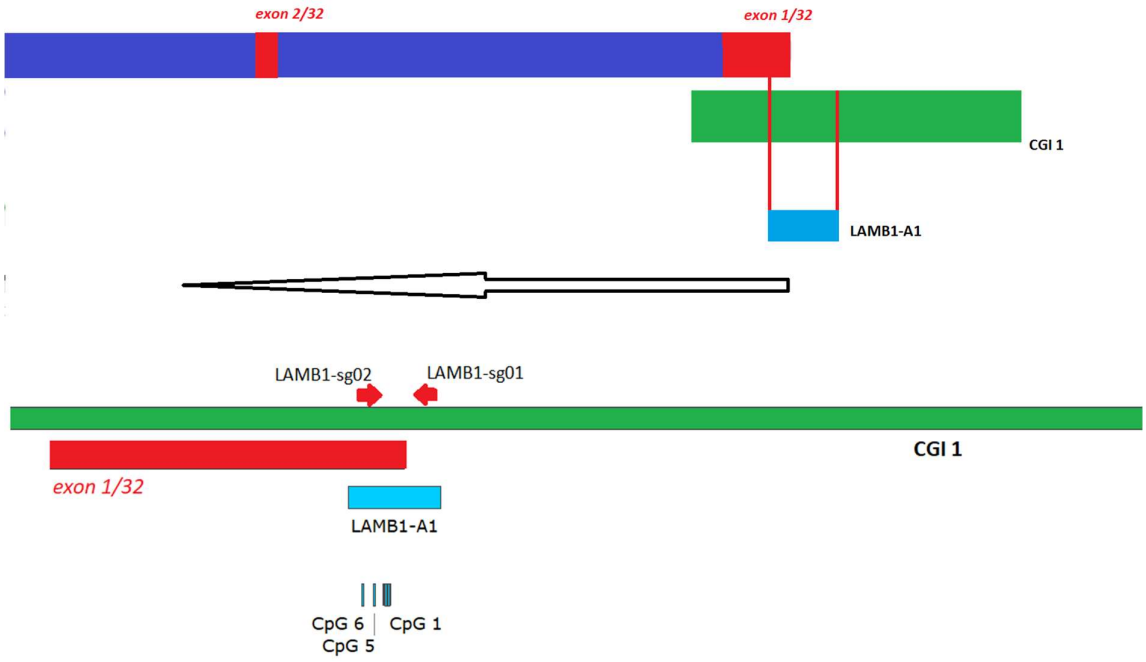


Figure 10. *LAMB1* gene location, scheme and sequence assayed for CpG methylation. **A)** *LAMB1* gene position on human chromosome 7 is marked with a red line on the chromosome scheme (q31.1). Representation taken from *www.genecards.org*., bands

according to *Ensembl* **B)** *LAMB1* gene is shown with its 32 exons (red boxes) and 31 introns (blue lines), with CGI in its promoter region, as represented in GRCh38/hg38. **C)** Assay LAMB1-A1 is located within this CGI, stretching through the initial part of the first exon (as indicated in the enlarged picture below, arrow indicating the transcript orientation). In GRCh38.p12 primary assembly, this CGI stretches from chr7 108001351 – 108003613 (-). Note that *LAMB1* gene is orientated on (-) strand and it is represented in that way in these schemes. In the additionally enlarged picture, assay LAMB1-A1 with 6 analysed CpG sites (blue lines, CpG 1-CpG 6) is shown with indicated positions of 2 sgRNAs chosen for dCas9-TET1 targeting (LAMB1-sg01 and LAMB1-sg02; red arrows indicating their position and orientation).

3.4 Cloning of sgRNAs into dCas9-TET1 constructs

All sgRNAs were synthesized as single stranded oligonucleotides (sense and antisense) by Macrogen Oligo DNA synthesis service. Additional G was added at 5' end to chosen guide sequences not containing it originally to enhance future transcription from U6 promoter (141), and 5' overhangs were added to make annealed oligonucleotides compatible with future cloning into expression vector via BbsI restriction sites. Oligonucleotides were annealed and phosphorylated in 1x T4 DNA ligase Buffer (Takara) with 5U of T4 PNK (NEB) in following conditions: 37°C for 30 min; 95°C for 5 min, followed by ramp-down cooling to 25°C. Annealed oligonucleotides were diluted and cloned into constructed dCas9-TET1 (and dCas9-DED1) tools using BbsI (BpiI) (10U, Thermo Scientific) and T4 DNA ligase (350U, Takara) in one-step digestion-ligation in 1x Buffer G (NEB), in 6 cycles of 37°C for 5 min and 23°C for 5 min. Ligation reactions were treated with RecBCD (5U, NEB) and 2 of each ligation reaction was transformed into XL10GOLD chemically competent *E. coli* strain (Agilent Technologies). Successful insertion of sgRNAs into dCas9-TET1 and dCas9-DED1 constructs was confirmed by Sanger sequencing with U6-Fw sequencing primer; 5'-GAGGGCCTATTTCCCATGATTCC-3'. Sequences of cloned sgRNAs are listed in Table 3.

Table 3. sgRNAs cloned into dCas9-TET1(DED1) constructs

Name	Sequence (5' → 3'), including PAM (underlined)	Target
MGAT3-sg01	CATTCGCTGGGATATAGAATAGG	*MGAT3* promoter (within CGI)
MGAT3-sg02	CCCTGCACCTTCGACGATGGCGG	*MGAT3* (within CGI)
MGAT3-sg03	ATGCCTTGTGCCTCGCACCGCGG	*MGAT3* (within CGI)
MGAT3-sg04	CTGCTCTGTAGGCCCCAGAACGG	*MGAT3* (within CGI)
MGAT3-sg05	GGACGCCTCTGAGCCCTGAGAGG	*MGAT3* (within CGI)
MGAT3-sg07	CGGCACCGTGCACACATCACAGG	*MGAT3* (within CGI)
MGAT3-sg08	AATCCCGGCCCAGGTTCACGCGG	*MGAT3* (within CGI)
LAMB1-sg01	ACACATCCACCCTTTGTTGGGGG	*LAMB1* promoter (within CGI)
LAMB1-sg02	AGCAGCGAGAGCCTCCCTCCCGG	*LAMB1* exon 1/32 (within CGI)
NT	GTAGGCGCGCCGCTCTCTAC	none

3.5 Cell culture conditions and transfection

Human embryonic kidney cells HEK293 were maintained in Dulbecco's Modified Eagle Medium (Sigma-Aldrich) supplemented with 10% heat inactivated fetal bovine serum (Sigma-Aldrich), 4 mM L–glutamine (Sigma-Aldrich), 100 U/ml penicillin and 100 µg/ml streptomycin (Sigma-Aldrich). Cells were cultured at 37°C in a 5% CO_2 humidified incubator.

All transient transfections were done with Lipofectamine 3000 Reagent (Invitrogen), either in 24-well plates or 6-well plates, according to manufacturer's instructions. OptiMEM (Thermo Fisher Scientific) was used for preparation of transfection mixtures.

3.6 Titration of plasmid DNA for subsequent experiments

Cells were seeded into a 24–well plate and transfected the next day at 70% confluence using Lipofectamine 3000 (Invitrogen), 1.5 µl of Lipofectamine and 1 µl of P3000 Reagent were used for each well. Transfections were done with either 20 ng, 30 ng, 40 ng, 50 ng, 75 ng or 100 ng of plasmids co–expressing dCas9-TET1-PuroR and sgRNA LAMB1-sg01 (Table 3), in biological duplicates. Control groups were cells transfected with dCas9-DED1-PuroR co–expressing sgRNA LAMB1-sg01, cells transfected with dCas9-TET1-puroR and the non-targeting sgRNA (Table 3) and mock-transfected cells (underwent transfection procedure without plasmid DNA).

One day after transfection, cells were passaged to 6–well plates and selected for 48 h with 1.5 µg/ml puromycin. Eight days after transfection, cells were harvested and incubated at 37°C overnight in lysis buffer (50 mM Tris pH 8.5, 1 mM EDTA, 0.5% Tween 20) with proteinase K (400 µg/ml) yielding crude cell lysates.

DNA from cell lysates was bisulfite converted and purified using EZ DNA Methylation-Gold Kit (Zymo Research Europe) according to the manufacturer's protocol. Briefly, DNA was mixed with CT Conversion Reagent, followed by 10 min of denaturation at 98°C, and 2.5 hours of incubation at 64°C, followed by desulphonization, washing and elution. Fragments of interest were amplified from bisulfite converted DNA with PyroMark PCR kit (Qiagen) according to the manufacturer's instructions. The cycling was performed as follows: initial denaturation for 15 min at 95°C; 50 cycles of 30 s at 95°C, 30 s at 54°C (MGAT3–A1 fragment), 60°C (MGAT3–A2 fragment) or 64°C (LAMB1-A1 fragment) and 30 s at 72°C; final extension for 10 min at 72°C. PCR amplicons were sequenced using the PyroMark Q24 Advanced pyrosequencing system (Qiagen). Sequences of PCR primers and pyrosequencing primers are listed in Table 4.

Table 4. Sequences of the PCR primers and pyrosequencing primers

Primer name	Sequence (5' → 3')	Use
MGAT3-A1-Fw	GTTGGGATATAGAATAGGTAG	PCR amplification of MGAT3-A1 fragment, forward primer was used for pyrosequencing
MGAT3-A1-Rev	[Btn]ACCATTCCTCTCAAAACTCA	
MGAT3-A2-Fw	GTTTTTGAGTTTTGAGAGGAATGG	PCR amplification of MGAT3-A2 fragment, forward primer was used for pyrosequencing
MGAT3-A2-Rev	[Btn]ACCCTCTTAAACCTACTCTCCTAC	
LAMB1-A1-Fw	GAAGTGGAGGGTTTAT	PCR amplification of LAMB1-A1 fragment
LAMB1-A1-Rev	[Btn]ATCAAATCTATCCAACAA	
LAMB1-A1-seq	TTGATTAGGGTGGG	Pyrosequencing of LAMB1-A1

3.7 Obtaining an activity profile of dCas9-TET1 molecular tool

40 000 cells per well were seeded into a 24–well plate and transfected the next day using Lipofectamine 3000 (Invitrogen), 1.5 µl of Lipofectamine and 1 µl of P3000 Reagent were used for each well. Transfections were done with 100 ng of plasmids co–expressing dCas9-TET1-PuroR and a chimeric sgRNA, in biological duplicates. All sgRNAs used in this experiment were targeting *MGAT3* promoter; MGAT3-sg01 – MGAT3-sg08 (Table 3, Figure 9). Control groups were cells transfected with either dCas9-DED1-PuroR co–expressing targeting sgRNAs, cells transfected with dCas9-TET1-puroR and the non-targeting sgRNA (Table 3) and mock-transfected cells.

One day after transfection, cells were passaged to 6–well plates and selected for 48 h with 1.5 µg/ml puromycin. Eight days after transfection, cells were harvested and incubated at 37°C overnight in lysis buffer with proteinase K as described before. Methylation of the MGAT3-A1 and MGAT3-A2 was assayed as described above.

3.8 Obtaining an activity profile of TET1-dCas9 molecular tool

TET1 catalytic domain prepared as described in section 3.2 was inserted N-terminally to dCas9 from *Streptococcus pyogenes* into a backbone for assembly of modular epitoolbox created in our lab (unpublished data) and TET1-dCas9 tool was generated (construct scheme shown in Figure 14). 40 000 cells per well were seeded into a 24–well plate and transfected the next day in the same way as described in section 3.7. Plasmids co–expressing TET1-dCas9 (TET1-dCas9-PuroR) and a chimeric sgRNA were used, in biological duplicates. All sgRNAs used in this experiment were targeting *MGAT3* promoter; MGAT3-sg01 – MGAT3-sg05 (Table 3, Figure 9). Control groups were cells transfected with either TET1-dCas9 co–expressing targeting sgRNAs, cells transfected with TET1-dCas9 and the non-targeting sgRNA (Table 3) and mock-transfected cells.

One day after transfection, cells were passaged to 6–well plates and selected for 48 h with puromycin. Eight days after transfection, cells were harvested and incubated at 37°C overnight in lysis buffer with proteinase K as described before. Methylation of the MGAT3-A1 and MGAT3-A2 was assayed as described above.

3.9 Time Course evaluation of dCas9-TET1 induced effects on *MGAT3* and *LAMB1* promoter methylation

500 000 cells per well were seeded into 6-well plates 24 h prior to transfection with 800 ng of dCas9-TET1-PuroR plasmids encoding MGAT3-sg03 or LAMB1-sg01 (Table 3) and Lipofectamine 3000 (Invitrogen), 7.5 µl of Lipofectamine and 5 µl of P3000 Reagent were used for each well in triplicates. Control groups were cells transfected with either dCas9-DED1-PuroR co–expressing targeting sgRNAs and mock-transfected cells. One day after transfection, triplicates were pooled in total 15 ml suspensions from which 1 ml was re-plated per well in new 6-well plates. Cells were selected as described above. Aliquots of the transfected cells were collected daily until the 8th day and after that on 10th, 13th, 15th, 20th, 24th and 30th day. On the 9th day cells reached 80-90% confluence and were collected and re-plated in p10 plates. On the 15th day cells reached 90% confluence and were collected, diluted 1:5 and re-plated in new p10 plates, same was done on 24th day. Harvested cells were incubated at 37°C overnight in lysis buffer with proteinase K as described before. Methylation

of the MGAT3-A1, MGAT3-A2 and LAMB1-A1 for each of the 14 time points was assayed as described above.

3.10 Analysis of hydroxymethylation level

500 000 cells per well were seeded into a 6–well plate and transfected the next day using Lipofectamine 3000 (Invitrogen), 1.5 µl of Lipofectamine and 1 µl of P3000 Reagent were used for each well. Transfections were done with 800 ng of plasmids co–expressing dCas9-TET1-PuroR and sgRNA MGAT3-sg01, plasmids co–expressing dCas9-DED1-PuroR and sgRNA MGAT3-sg01, plasmids co–expressing dCas9-TET1-PuroR and non-targeting sgRNA in biological triplicates. Control group were mock-transfected cells. Cells were selected for 48 h with puromycin as described before.

8th day after transfection cells were collected from wells and DNA was isolated with DNeasy Blood&Tissue Kit (Qiagen), according to manufacturer's protocol. Isolated DNA was selectively chemically oxidized to achieve 5-hydroxymethylcytosine (5hmC) to 5-formylcytosine (5fC) conversion which enables bisulfite conversion of 5fC to uracil. General denaturation and oxidation protocol was adapted from previously published work on hydroxymethylation level analysis, following instructions for genomic DNA manipulation (122).

Briefly, one microgram of each isolated DNA sample was denatured with 0.05 M NaOH in total 24 µl reaction followed by oxidation with 1 µl of 15 mM potassium perruthenate ($KRuO_4$). Total 25 µl of each oxidation reaction was purified using Mini Quick Spin Columns (Roche). Oxidized DNA was bisulfite converted, amplified and methylation of the MGAT3-A1 and MGAT3-A2 was assayed by bisulfite pyrosequencing as described above. Same procedure (bisulfite conversion, amplification and pyrosequencing) was done with the same amount of unoxidized isolated DNA for each sample and methylation levels of oxidized and unoxidized samples were subtracted to gain information about hydroxymethylation levels.

3.11 Targeting the *MGAT3* promoter with dCas9-TET1 and pooled sgRNAs

40 000 cells per well were seeded into a 24–well plate and transfected the next day using Lipofectamine 3000 (Invitrogen), 1.5 µl of Lipofectamine and 1 µl of P3000 Reagent were used for each well. Transfections were done with 100 ng of equimolar mixtures of five plasmids co–expressing dCas9-TET1-PuroR and a chimeric sgRNA, in biological triplicates. All sgRNAs used in this experiment were targeting the *MGAT3* promoter; MGAT3-sg01 – MGAT3-sg05 (Table 3). Control groups were cells transfected with either equimolar mixtures of five dCas9-DED1-PuroR plasmids co–expressing targeting sgRNAs, cells transfected with dCas9-TET1-PuroR and the non-targeting sgRNA (Table 3) and mock-transfected cells.

One day after transfection, cells were passaged to 6–well plates and selected for 48 h with 1.5 µg/ml puromycin. Eight days after transfection, cells were collected and DNA and RNA were isolated with RNA/DNA Purification Kit (Norgen Biotek). DNA was bisulfite converted, amplified and methylation of the MGAT3-A1 and MGAT3-A2 was assayed as described above.

3.12 Gene expression analysis

One microgram of total isolated RNA was treated with TURBO DNase (2U, Invitrogen) and used as a template for reverse transcription using the PrimeScript RTase (100U, Takara) and random hexamer primers (Invitrogen) with use of Recombinant RNase Inhibitor (20U, Takara). Quantitative real-time PCR (RT–qPCR) was performed on cDNA from cells transfected with pooled sgRNAs co–expressing dCas9-TET1-PuroR and all three control groups. RT–qPCR was done with TaqMan hydrolysis probes in the 7500 Fast Real-Time PCR system (Applied Biosystems), in technical duplicates with TaqMan Gene Expression Master Mix and the following TaqMan Gene Expression Assays: Hs02379589_s1 (*MGAT3*), Hs01369240_m1 (*MGAT3*), Hs02800695_m1 (*HPRT1*) and Hs00609297_m1 (*HMBS*) (Applied Biosystems). Hs02379589_s1 and Hs01369240_m1 expression assays were used for *MGAT3* as they differ in detection; Hs02379589_s1 detects two transcripts of *MGAT3* gene but has a drawback as it does not span exons and detects genomic DNA as well. Hs01369240_m1 spans exons, but detects only the longer transcript. The expression of the

target gene *MGAT3* was normalized to the reference genes *HPRT1* and *HMBS* and analysed using the comparative ddCt method (142).

3.13 Targeting the *MGAT3* gene with VPR-dCas9 and pooled sgRNAs

For VPR-dCas9 mediated activation of *MGAT3* a construct containing an N-terminal fusion of VPR activation domain with catalytically inactive Cas9 from *Staphylococcus aureus* (SaCas9) previously constructed in our lab (data not shown) was used. Construct scheme is shown in Figure 14. Different sgRNAs were designed for VPR-dCas9 (VPR-d(Sa)Cas9-PuroR) targeting because of two important differences in the approach; (i) recommended targeting regions for VPR targeted activation are short stretches of bases upstream of predicted TSSs, (ii) sgRNAs had to meet design requirements for SaCas9. SaCas9 requires a 5'-NNGRRT-3' PAM immediately downstream of the target sequence. Positions of chosen sgRNAs are shown in Figure 11. Designed sgRNAs were cloned into VPR-dCas9 construct as described in section 3.4. Sequences of cloned sgRNAs are shown in Table 5.

500 000 cells per well were seeded into 6-well plates 24 h prior to transfection with 1 µg of equimolar mixture of three plasmids co–expressing VPR-dCas9 construct and a chimeric sgRNA, in biological duplicate. Transfections were also done with 1 µg of each VPR-dCas9 construct expressing a single sgRNA, in duplicates. Control group were mock-transfected cells. Cells were selected as described before, and total RNA was isolated 5 days after transfection using RNeasy Mini Kit (Qiagen). RNA was treated with TURBO DNase (Invitrogen) and reverse transcribed as described earlier. RT–qPCR was performed on cDNA from cells transfected with pooled sgRNAs co–expressing VPR-dCas9, cells transfected with each single sgRNA co-expressed with VPR-dCas9 and mock-transfected cells. RT–qPCR was done with TaqMan hydrolysis probes, in technical duplicates with TaqMan Gene Expression Master Mix and the following TaqMan Gene Expression Assays: Hs02379589_s1 (*MGAT3*), Hs01369240_m1 (*MGAT3*) and Hs00609297_m1 (*HMBS*) (Applied Biosystems). The expression of the target gene *MGAT3* was normalized to the reference gene *HMBS* and analysed using the comparative ddCt method (142).

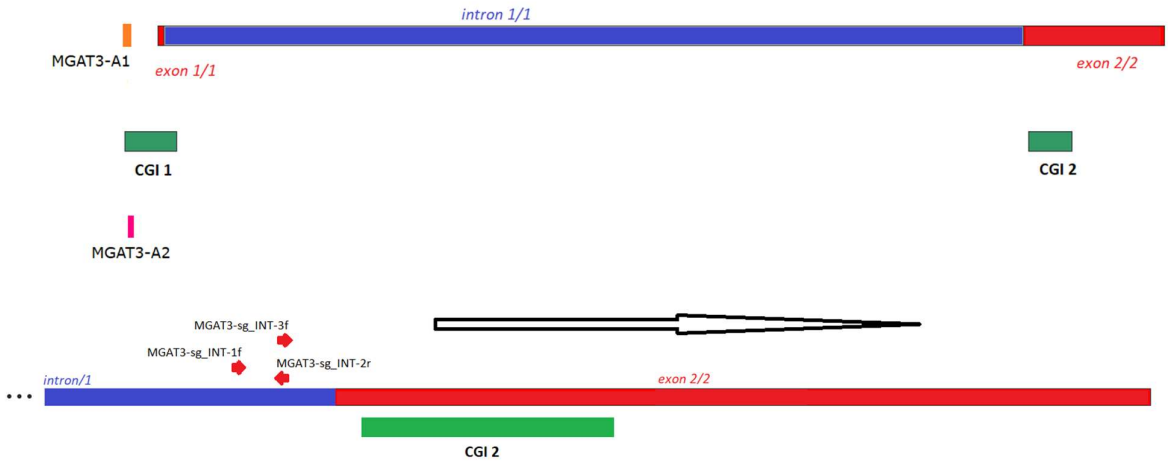


Figure 11. Positions of sgRNAs for VPR-dCas9 targeting of the *MGAT3* gene. In the first picture, the *MGAT3* gene scheme is shown with its 2 exons (red boxes) and 1 intron (blue box) and 2 CGIs, as represented in GRCh38/hg38. Assays MGAT3-A1 and MGAT3-A2 are located immediately upstream and in the beginning of the first CGI (indicated by orange and pink lines). In the enlarged picture below, sgRNAs chosen for VPR-dCas9 activation of *MGAT3,* MGAT3-sg_INT-1f, MGAT3-sg_INT-2r and MGAT3-sg_INT-3f, are shown in alignment with *MGAT3* gene and CGI 2 sequence (sgRNA positions and orientations are indicated with red arrows). Black arrow represents *MGAT3* transcript orientation.

Table 5. sgRNAs cloned into VPR-dCas9 constructs

Name	**Sequence (5' → 3'), including PAM (underlined)**	**Target**
MGAT3-sg_INT-1f	CCTCAGCCTCTGAAGGTGTTGGGAT	*MGAT3* – intron 1/1
MGAT3-sg_INT-2r	CACTGCTAGGCTTGGCTCCAGGAGT	
MGAT3-sg_INT-3f	TGCAGCCGCACAGTCAGGGTGGGGT	

3.14 Data analysis and statistics

Sequence analysis:

Sequencing data was analysed with SnapGene software (GSL Biotech, Chicago, Illinois, USA).

Plasmid maps and pyrosequencing assays schemes were created with SnapGene software.

Methylation data analysis:

Data is shown as arithmetic means of two or three biological replicas. Standard deviations between replicas are indicated in each figure.

Gene expression data analysis:

Statistical analyses were done with SPSS software (IBM Analytics, Armonk, New York, USA). Different biological groups were compared among each other using a one-way analysis of variance (ANOVA) followed by *post hoc* Tukey test of multiple comparisons. P values of ≤ 0.05 were considered statistically significant.

4. Results

4.1 Construction of dCas9-TET1 molecular tool

I have developed the molecular tool dCas9-TET1 for targeted DNA demethylation in human cells. It is a fusion of the human codon optimized *S. pyogenes* dCas9 (4101 bp, 1367 aa) and the human TET1 CD (2163 bp, 721 aa), which retains the original TET1 catalytic activity. dCas9 was created by introducing the H841A mutation to the HNH nuclease domain of human codon optimized *S.pyogenes* nickase Cas9n already containing the D10A silencing mutation in RuvC nuclease domain in plasmid pSpCas9(BB)-2A-Puro (plasmid #48141, Addgene plasmid repository) (11). Human TET1 CD was PCR amplified from plasmid pJFA344C7 (plasmid #49236, Addgene plasmid repository), which is an expression vector for TALE-derived DNA binding domain fused C-terminally via Gly_4Ser linker to synthesized minimal TET1 CD expression driven by EF1a promoter, successfully used for targeted CpG demethylation and gene activation (69). TET1 CD is fused to C-terminus of dCas9 via Gly_4Ser linker, one of the most commonly used flexible linkers which offers fused domains a certain level of movement favourable for desired interactions (143). Catalytically inactive version of TET1 catalytic domain named DED1 variant was created by introducing two substitutions, H1671Y, D1673A, known to abolish enzymatic activity of TET1 by making its active site inaccessible for iron ion binding (21).

T2A-PuroR (Figure 12A) and T2A-eGFP (Figure 12B) fragments were inserted C-terminally of TET1(DED1) CD in dCas9-TET1(DED1) fusion for successful manipulation of transfected cells – either puromycin selection, or visualization of transfected cells. EcoRI fragment containing T2A-eGFP was isolated from plasmid pSpCas9n(BB)-2A-GFP (plasmid #48140, Addgene plasmid repository) (11). Mutation H166R was introduced into puromycin resistance gene of pSpCas9n(BB)-2A-Puro (plasmid #48141, Addgene plasmid repository), to achieve improved puromycin selection like the one in plasmid pSpCas9n(BB)-2A-Puro V2.0 (plasmid #62987, Addgene plasmid repository) (11), and EcoRI fragment encoding T2A-Puro_H166R was isolated from selected successfully mutated pSpCas9n(BB)-2A-Puro_H166R clone. T2A is a short self-cleaving peptide isolated from *thosea asigna* virus, it mediates a *ribosome skipping* effect, co-translational cleavage of polyproteins transcribed from the same promoter as a single transcript (144). Fusion protein scheme is

shown in Figure 13. Polypeptide sequences of constructed fusion proteins are listed in Appendix 2.

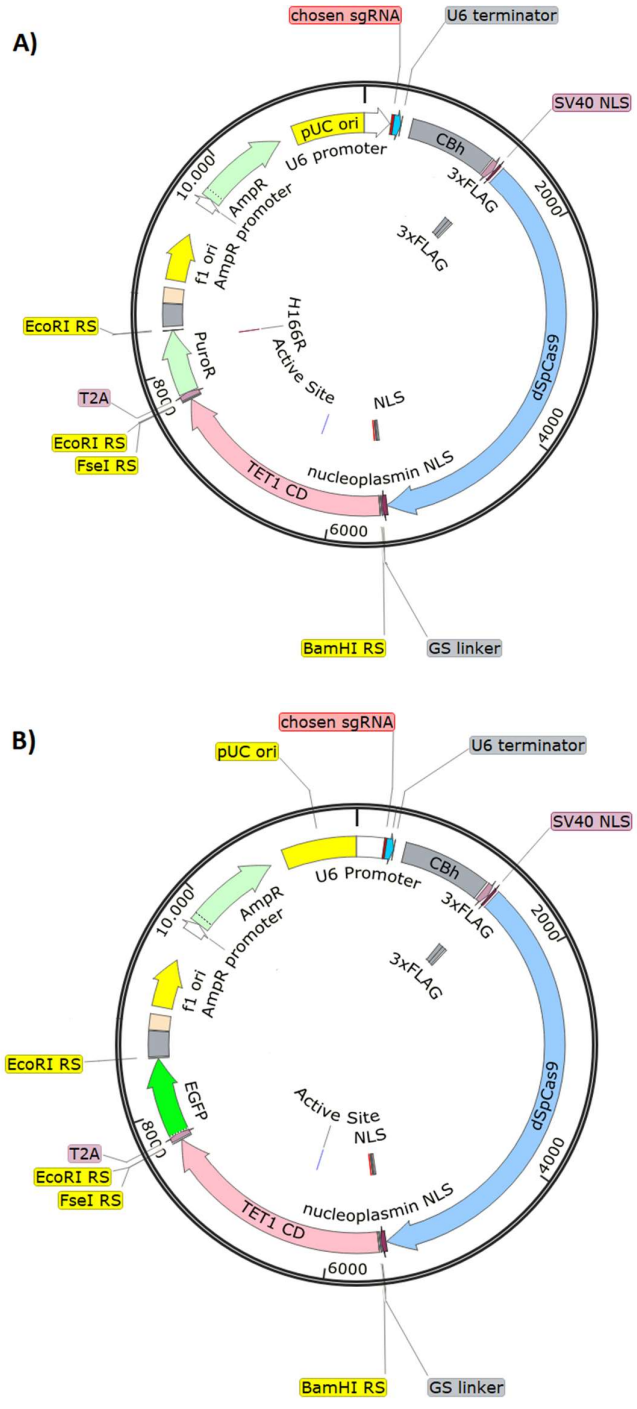


Figure 12. Maps of expression vectors for created dCas9-TET1 molecular tools and chosen sgRNA molecules. A) Map of dCas9-TET1-PuroR and sgRNA expression plasmid.

Total size of dCas9-TET1-PuroR expression plasmid is 11 373 bp, and it contains several distinctive features. Expression of Cas9 from *S. pyogenes* containing inactivating mutations in RuvC and HNH cleavage domains (D10A; H841A) (dSpCas9) fused with catalytic domain of human TET1 (TET1 CD) via short Gly-Gly-Gly-Gly-Ser (GS) linker is driven by chicken β-actin (CBh) promoter. Restriction enzyme sites used for TET1 CD insertion are indicated (BamHI and FseI RSs). Nucleotide sequence of CBh is 799 bp long, dSpCas9 is 4101 bp, short GS linker is 15 bp, and TET1 CD is 2163 bp. TET1 CD Active Site is marked, which is mutated (H1671Y; D1673A) in catalytically dead construct variant dCas9-DED1-PuroR. Resistance marker for puromycin (PuroR, 600 bp) is cloned C-terminally of TET1 CD, downstream of self-cleaving T2A (54 bp) peptide. Site of puromycin resistance gene mutation H166R is indicated. EcoRI restriction sites used for cloning of PuroR are highlighted in yellow (EcoRI RSs). There is an epitope tag, 3xFLAG sequence of 66 bp, downstream of CBh promoter, immediately followed by SV40 nuclear localization signal (SV40 NLS, 21 bp) and a nucleoplasmin NLS of 48 bp is cloned immediately downstream of dSpCas9. Plasmid backbone contains bacterial origin of replication (ori) and a resistance marker for ampicillin (AmpR, 861 bp) under its own promoter (AmpR promoter, 105 bp). Expression of cloned sgRNA followed by sgRNA scaffolds of 766 bp is driven by human U6 polymerase III promoter (U6 promoter, 249 bp), this transcript ends with U6 terminator (66 bp). **B)** Map of dCas9-TET1-eGFP and sgRNA expression plasmid. Total size of dCas9-TET1-eGFP expression plasmid is 11 484 bp, and it contains the same described features as dCas9-TET1-PuroR and sgRNA expression vector, except for the eGFP selection marker (714 bp) instead of PuroR.

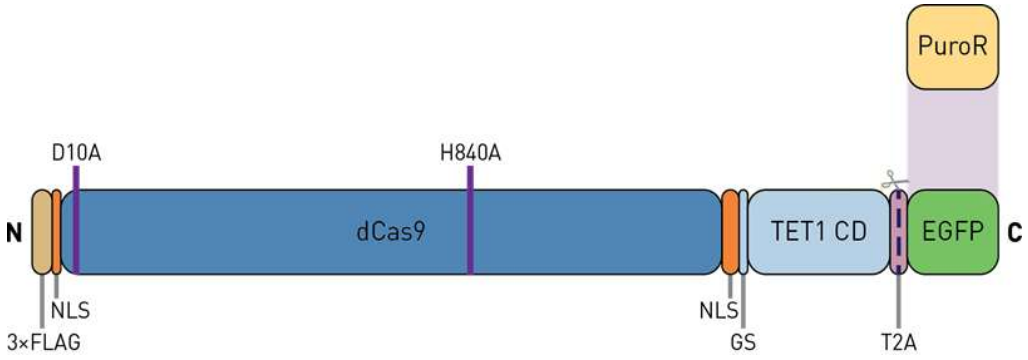


Figure 13. Domains of the fusion protein dCas9-TET1. Silencing substitutions (D10A and H840A) in *Streptococcus pyogenes* Cas9 nuclease domains are indicated. Catalytic domain of

human TET1 (TET1 CD) is fused C-terminally to dCas9 via short Gly_4Ser (GS) linker. dCas9-TET1 is expressed as a bicistronic mRNA, along with puromycin resistance gene or eGFP (PuroR/EGFP), allowing selection of transfected cells. PuroR/EGFP module is efficiently released after translation due to a self-cleaving 2A peptide (T2A) from *Thosea asigna virus 2A*. Nuclear localization signals (NLSs) are located at both ends (N- and C-terminus) of the dCas9 part of the fusion protein, while the fusion domains are connected with a flexible Gly_4Ser linker (GS).

4.2 Insertion of TET1 and VPR into the modular epitoolbox backbone

TET1 catalytic domain (TET1 CD) prepared as described in section 3.2 was inserted N-terminally to dCas9 from *Streptococcus pyogenes* (TET1-dCas9) into a backbone for assembly of modular epitoolbox created in our lab (unpublished data).

The VPR-dCas9 construct used for VPR-mediated activation of *MGAT3* was generated by insertion of VPR activation domain into the same modular epitoolbox backbone N-terminally to dCas9 from *Staphylococcus aureus.* Construct schemes are shown in Figure 14.

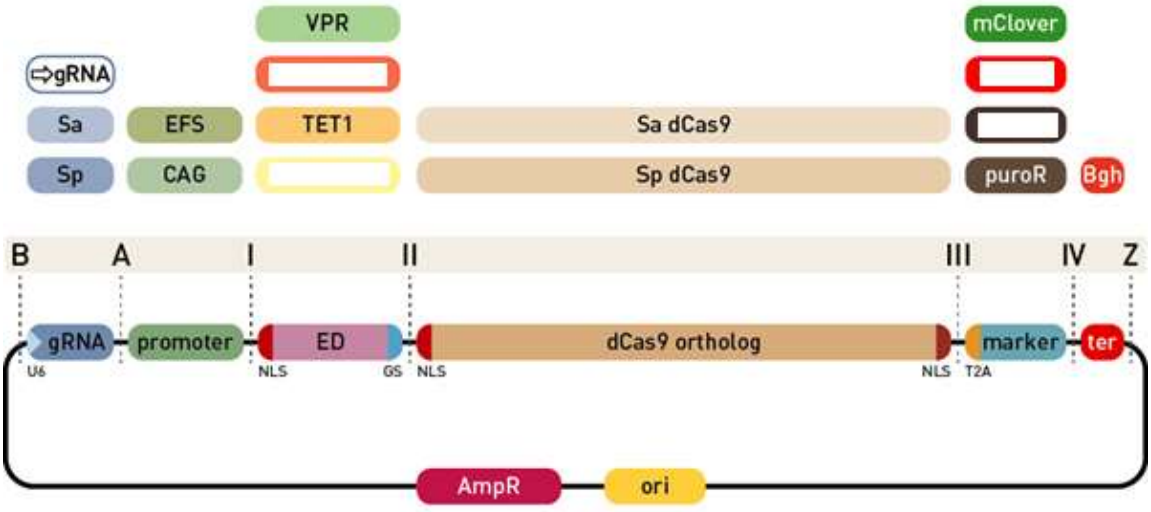


Figure 14. Backbone of the modular epitoolbox for N-terminal dCas9 fusions. TET1 CD and VPR activation domain were inserted into a plasmid backbone along with other specific

modules using Golden Gate cloning strategy. Each position for insertion of modules into the backbone has ends compatible with certain type of module ends. The first position ("B" to "A") receives a gRNA expression module with SaCas9 or SpCas9 scaffold, an empty module for gRNA cloning with red-white selection, or a multi-guide module for a second step insertion of up to 6 gRNAs. Position "A" to "I" is for insertion of a eukaryotic promoter, followed by the effector domain ("I" to "II") containing an N-terminal NLS and a short G_4S linker to a dCas9 ortholog ("II" to "III"), followed by a selection marker (fluorescence or antibiotic resistance, "III" to "IV") linked via the self-cleaving T2A peptide, which can be substituted with the dual-marker system. Finally, a module for eukaryotic transcription terminator is inserted between ends "IV" and "Z". The TET1-dCas9 construct contains the following modules: SpCas9 compatible sgRNA (one *MGAT3* targeting sgRNA inside position "B" to "A"; constructs with sgRNAs MGAT3-sg01-sg05 were created), CAG promoter ("A" to "I"), TET1 CD ("I" to "II"), dSpCas9 ("II" to "III"), puromycin resistance marker ("III" to "IV") and a Bgh transcription terminator ("IV" to "Z"). The VPR-dCas9 construct contains the following modules: SaCas9 compatible sgRNA (one *MGAT3* targeting sgRNA inside position "B" to "A"; constructs with sgRNAs sg_INT-1f, sgRNA_INT-2r and sgRNA_INT3f were created), CAG promoter ("A" to "I"), VPR activation domain ("I" to "II"), dSaCas9 ("II" to "III"), puromycin resistance marker ("III" to "IV") and a Bgh transcription terminator ("IV" to "Z"). Blank modules indicated above position "I" to "II" can be filled with any CD of choice and inserted into the backbone, and blank modules indicated above position "III" to "IV" can be filled with any selection marker of choice and inserted into backbone. NLS = nuclear localization signal; EFS = short EF1alpha promoter, CAG = cytomegalovirus (CMV) enhancer fused to the chicken beta-actin promoter; AmpR = ampicillin resistance marker; ori = origin of replication.

4.3 Titration of plasmid DNA for subsequent experiments

In the initial transfections of HEK293 cells with dCas9-TET1 molecular tool, the optimal amount of plasmid DNA was determined for subsequent experiments. On-target and off-target effect, as well as transfection efficiency were analysed. LAMB1-sg01 was used for this optimization. Methylation status of 6 CpG sites in LAMB1-A1 assay was determined by bisulfite pyrosequencing for the following experimental groups: i) cells transfected with either

20 ng, 30 ng, 40 ng, 50 ng, 75 ng or 100 ng of dCas9-TET1-PuroR and LAMB1-sg01 expression plasmid; ii) cells transfected with the same amounts of dCas9-DED1-PuroR and LAMB1-sg01 expression plasmid; iii) cells transfected with the same amounts of dCas9-TET1-PuroR and non-targeting (NT) sgRNA expression plasmid and iv) mock transfected cells (Figure 15). In subsequent experiments, 100 ng of plasmid DNA was used in 24-well plates and 800 ng in 6-well plates (data from optimization experiments not shown).

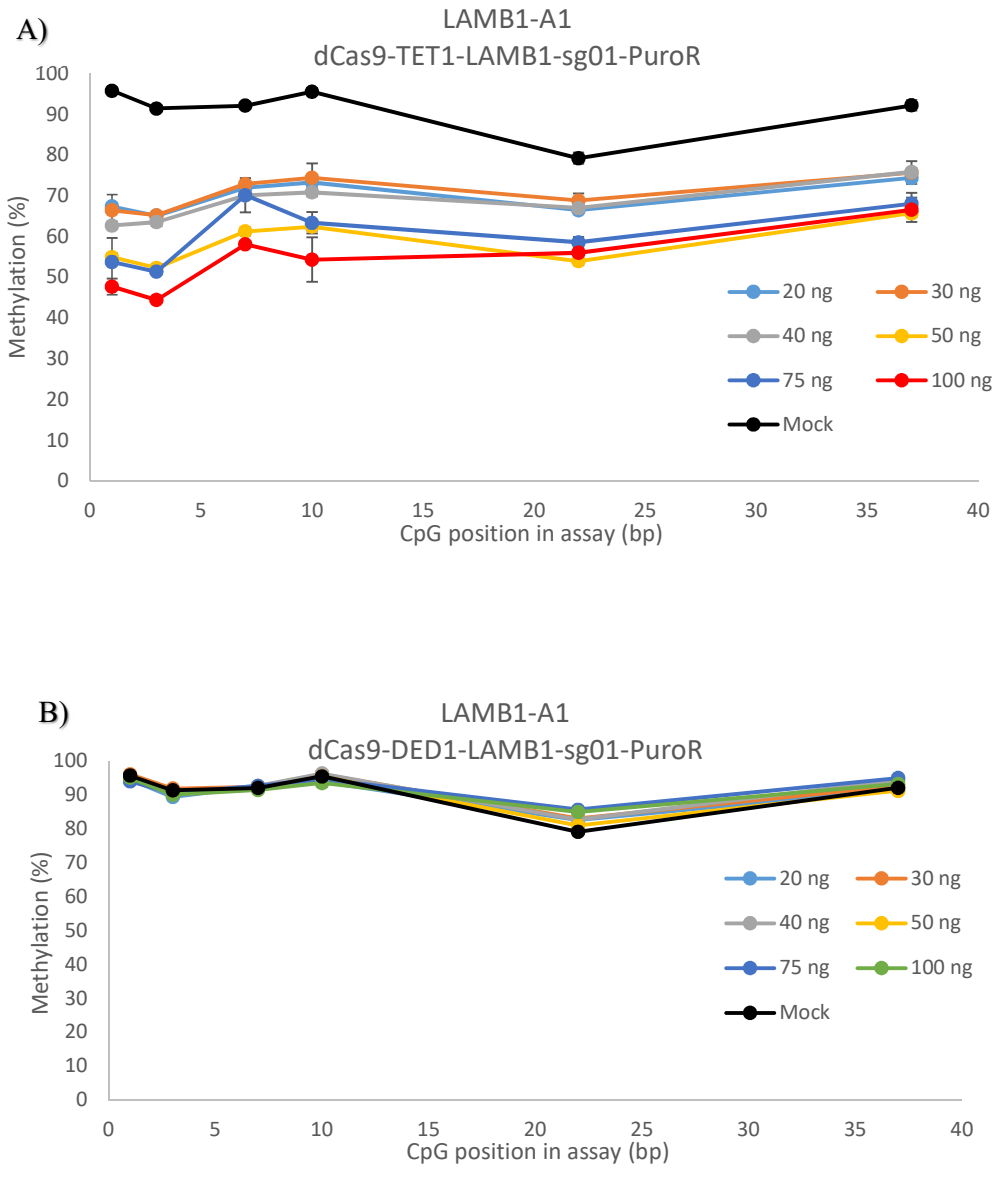


Figure 15. Figure is continued and described on the following page.

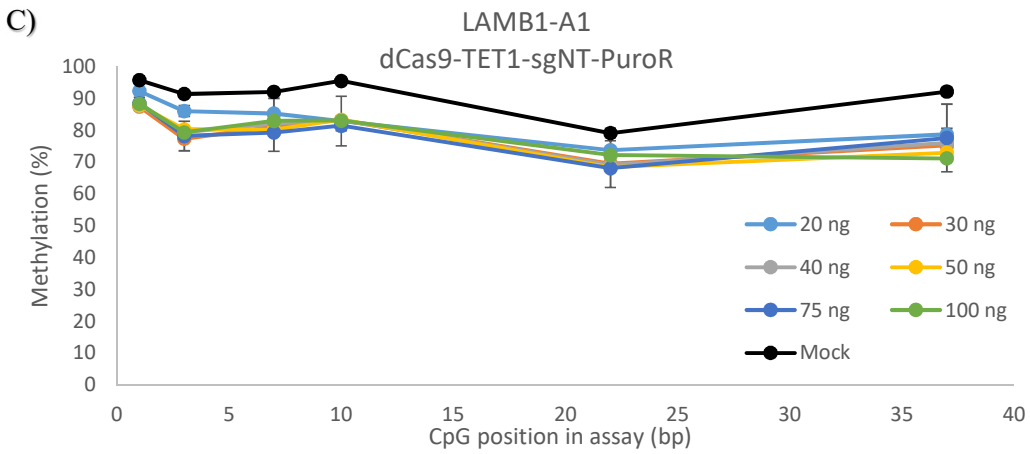


Figure 15. Titration of dCas9-TET1-PuroR and sgRNA expression plasmid DNA for subsequent experiments. A) Methylation levels of individual CpG sites within LAMB1-A1 assay are shown for HEK293 cells transfected with tested amounts (20ng, 30ng, 40ng, 50ng, 75ng, 100ng) of the dCas9-TET1-LAMB1-sg01 expression plasmid as lines connecting individual CpG sites (presented as dots). Each dot represents an arithmetic mean of two replicas, with indicated standard deviation bars. Methylation levels of mock transfected cells are shown in comparison. Most drastic methylation change is achieved in cells transfected with 100 ng of plasmid DNA (48% and 47% at first two CpG sites). **B)** Methylation levels of individual CpG sites within LAMB1-A1 assay are shown for cells transfected with tested amounts (20ng – 100ng) of dCas9-DED1-LAMB1-sg01 expression plasmid as lines connecting individual CpG sites (presented as dots), in comparison with mock transfected cells. Each dot represents an arithmetic mean of two replicas, with indicated standard deviation bars. No significant methylation changes were observed for any of the tested DNA amounts. **C)** Methylation levels of individual CpG sites within LAMB1-A1 assay are shown for cells transfected with tested amounts (20ng – 100ng) of dCas9-TET1-sgNT as lines connecting individual CpG sites (presented as dots), in comparison with mock transfected cells. Each dot represents an arithmetic mean of two replicas, with indicated standard deviation bars. Significant off-target effect was observed for all tested DNA amounts. No significant difference in off-target activity was observed for cells transfected with 30 ng, 40 ng, 50 ng, 75 ng and 100 ng of plasmid DNA.

4.4 Activity profile of dCas9-TET1 molecular tool

Methylation levels on CpG sites covered by MGAT3-A1 and MGAT3-A2 assays were analysed in HEK293 cells: (i) transfected with plasmids co–expressing dCas9-TET1-PuroR using 7 sgRNAs which targeted *MGAT3* promoter; (ii) transfected with dCas9-DED1-PuroR co–expressing targeting sgRNAs, (iii) transfected with dCas9-TET1-PuroR and non-targeting (NT) sgRNA and mock transfected cells. dCas9-TET1 targeting the *MGAT3* promoter resulted in decrease of methylation level at CpG sites covered by assays MGAT3-A1 and MGAT3-A2 in comparison with both dCas9-TET1-sgNT and mock transfected cells.

Each sgRNA induced variable levels of demethylation at specific CpG sites, depending on the sgRNA binding site (Figure 16). Methylation levels of individual CpG sites between cells transfected with dCas9-TET1-sgNT did not significantly differ from those of mock transfected cells (+/- 0.7 – 5.5%) except for the first CpG site within MGAT3-A2 assay (18.5% methylation decrease for dCas9-TET1-sgNT transfected cells). dCas9-TET1-MGAT3-sg01 transfection resulted in peak demethylation activity of 48% at third CpG site of the MGAT3-A1 assay (22 bp downstream from PAM recognition sequence) and 51% at first CpG site of the MGAT3-A2 assay (186 bp downstream from PAM) in comparison with mock-transfected cells. MGAT3-sg02 guided demethylation resulted in peak demethylation activity of 51.9% at sixth CpG site of MGAT3-A1 (17 bp downstream from PAM recognition sequence) and 36.4% at first CpG site of MGAT3-A2 (155 bp downstream from PAM). dCas9-TET1-MGAT3-sg03 transfection resulted in peak demethylation activity of 22.7% at third CpG site of MGAT3-A1 (57 bp upstream from PAM) and 51% at first CpG site of the MGAT3-A2 assay (105 bp from PAM). dCas9-TET1-MGAT3-sg04 transfection resulted in peak demethylation activity of 18.8% at 9th CpG site of MGAT3-A1 (34 bp upstream from PAM) and 47.6% at first CpG site of the MGAT3-A2 assay (72 bp from PAM). MGAT3-sg05 guided demethylation resulted in peak demethylation activity of 68.3% at first CpG site of the MGAT3-A2 assay (20 bp from PAM) in comparison with mock-transfected cells. dCas9-TET1-MGAT3-sg07 and dCas9-TET1-MGAT3-sg08 transfected cells had no observable methylation changes for analysed CpG sites. Their respective binding sites are 236 bp and 296 bp downstream of the last analysed CpG site in the assay MGAT3-A2. These results indicate that distance above 230 bp surpasses the activity range of dCas9-TET1 molecular tool.

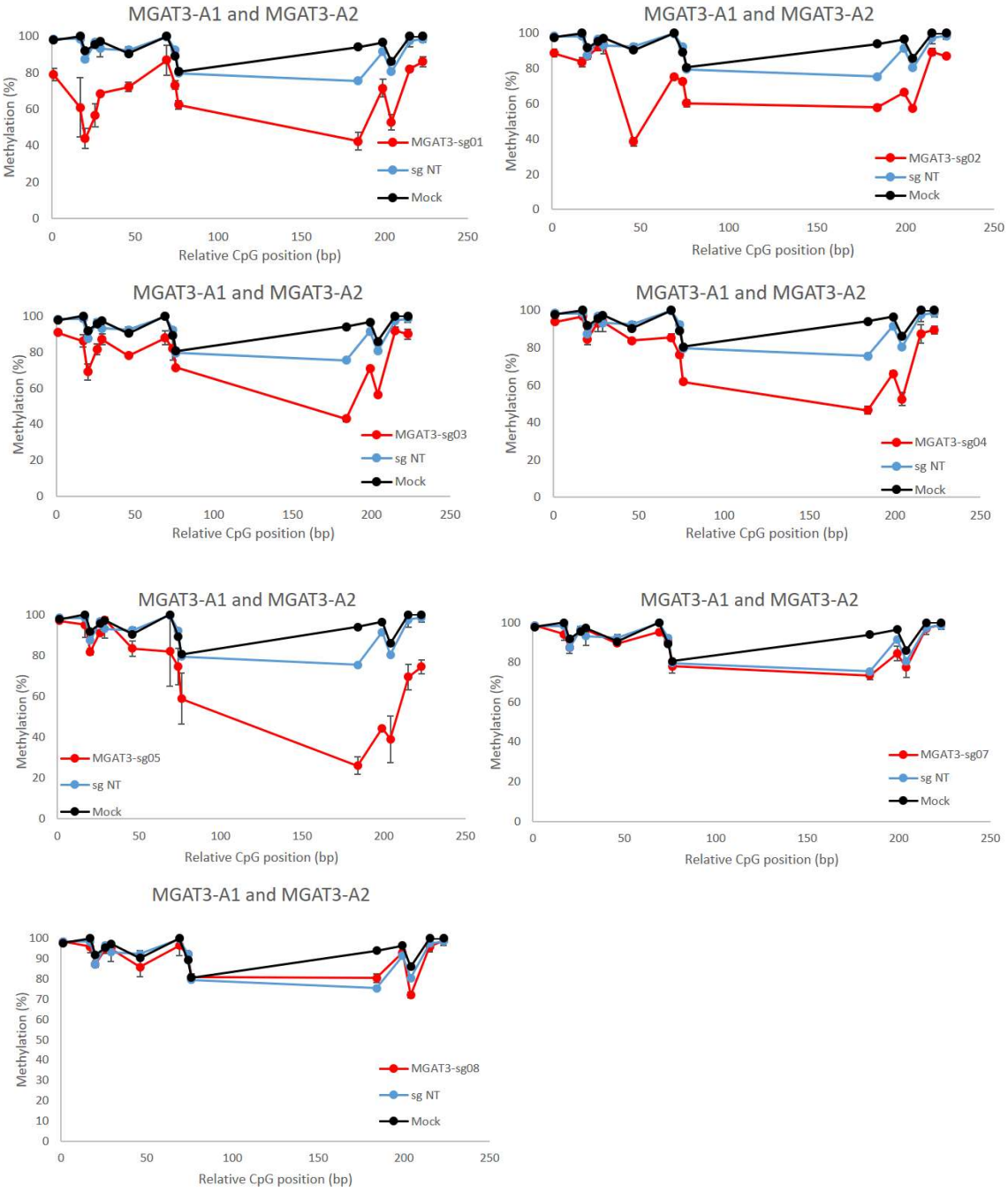


Figure 16. Targeted demethylation of the *MGAT3* promoter with dCas9-TET1 and 7 individual sgRNAs. Activity of dCas9-TET1 fusion protein, reflected on methylation level, guided by each individual sgRNA (MGAT3-sg01 – MGAT3-sg08) is shown as line connecting individual CpG sites (presented as dots). Same values for non-targeting control group (blue) and mock transfected cells (black) are shown in comparison to each sgRNA

(red). Methylation status of each individual CpG site within the assays MGAT3-A1 and MGAT3-A2 was determined by bisulfite pyrosequencing. Relative CpG position and distance is plotted on X-axis (the assays MGAT3-A1 and MGAT3-A2 with total coverage of 111 bp are shown together). Each dot represents an arithmetic mean of two replicas, with indicated standard deviation bars.

Summary demethylation activity profile of dCas9-TET1 was obtained from correlating upstream and downstream demethylation activity values for analysed CpG sites of MGAT3-A1 and MGAT3-A2 assays for each tested sgRNA with distance from sgRNA binding sequence, 1st base downstream of PAM sequence counted as base 0 (Figure 17).

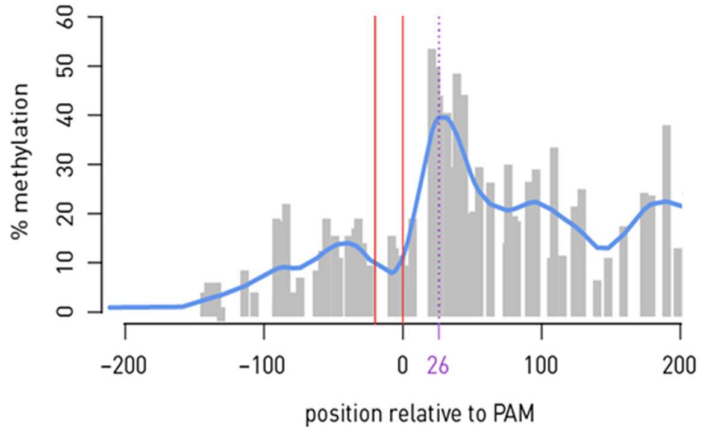


Figure 17. Summary profile of the dCas9–TET1 activity. Absolute methylation fraction decrease (compared to mock–transfected cells) is shown relative to the distance of each CpG site from the PAM sequence. Peak demethylation activity (purple dotted line) was observed 26 bp downstream from PAM recognition sequence (fraction indicated by red lines). Significant demethylation activity was also observed around 180-200 bp downstream of PAM sequence.

4.5 Activity profile of TET1-dCas9 molecular tool

Methylation levels on CpG sites covered by MGAT3-A1 and MGAT3-A2 assays were analysed in HEK293 cells: (i) transfected with plasmids co–expressing TET1-dCas9-PuroR using 5 sgRNAs which targeted *MGAT3* promoter; (ii) transfected with DED1-dCas9-PuroR co–expressing targeting sgRNAs, (iii) transfected with TET1-dCas9-PuroR and non-targeting (NT) sgRNA and mock transfected cells. TET1-dCas9 targeting the *MGAT3* promoter resulted in decrease of methylation level at CpG sites covered by assays MGAT3-A1 and MGAT3-A2 in comparison with DED1-dCas9 and targeting sgRNAs, TET1-dCas9-sgNT and mock transfected cells.

Each sgRNA induced variable levels of demethylation at specific CpG sites, depending on the sgRNA binding site (Figure 18). TET1-dCas9-MGAT3-sg01 transfection resulted in peak demethylation activity of 76.4% at third CpG site of the MGAT3-A1 assay (22 bp downstream from PAM recognition sequence) and 78.3% at first CpG site of the MGAT3-A2 assay (186 bp downstream from PAM) in comparison with mock-transfected cells. MGAT3-sg02 guided demethylation resulted in peak demethylation activity of 76.6% at eighth CpG site of MGAT3-A1 (45 bp downstream from PAM recognition sequence) and 78.3% at first CpG site of MGAT3-A2 (155 bp downstream from PAM). TET1-dCas9-MGAT3-sg03 transfection resulted in peak demethylation activity of 76.9% at first CpG site of the MGAT3-A1 assay (76 bp upstream from PAM) and 70.1% at first CpG site of the MGAT3-A2 assay (105 bp downstream from PAM). MGAT3-sg04 guided demethylation resulted in peak demethylation activity of 69.5% at eighth CpG site of the MGAT3-A1 assay (36 bp upstream from PAM) and 64.7% at first CpG site of the MGAT3-A2 assay (72 bp downstream from PAM). MGAT3-sg05 guided demethylation resulted in peak demethylation activity of 40.8% at 9th CpG site of MGAT3-A1 (86 bp upstream from PAM) and 87.1% at second CpG site of the MGAT3-A2 assay (20 bp downstream from PAM) in comparison with mock-transfected cells. Methylation levels of individual CpG sites between cells transfected with TET1-dCas9-sgNT did not significantly differ from those of mock transfected cells (+/- 0.7 – 5.5%) for MGAT3-A1 assay. TET1-dCas9-sgNT transfection resulted in 23.8%, 9.6% and 11.5% methylation decrease at first, second and third CpG site of MGAT3-A2, respectively.

Summary demethylation activity profile of TET1-dCas9 was obtained from correlating upstream and downstream demethylation activity values for analysed CpG sites of MGAT3-

A1 and MGAT3-A2 assays for each tested sgRNA with distance from sgRNA binding sequence, 1st base downstream of PAM sequence counted as base 0 (Figure 19).

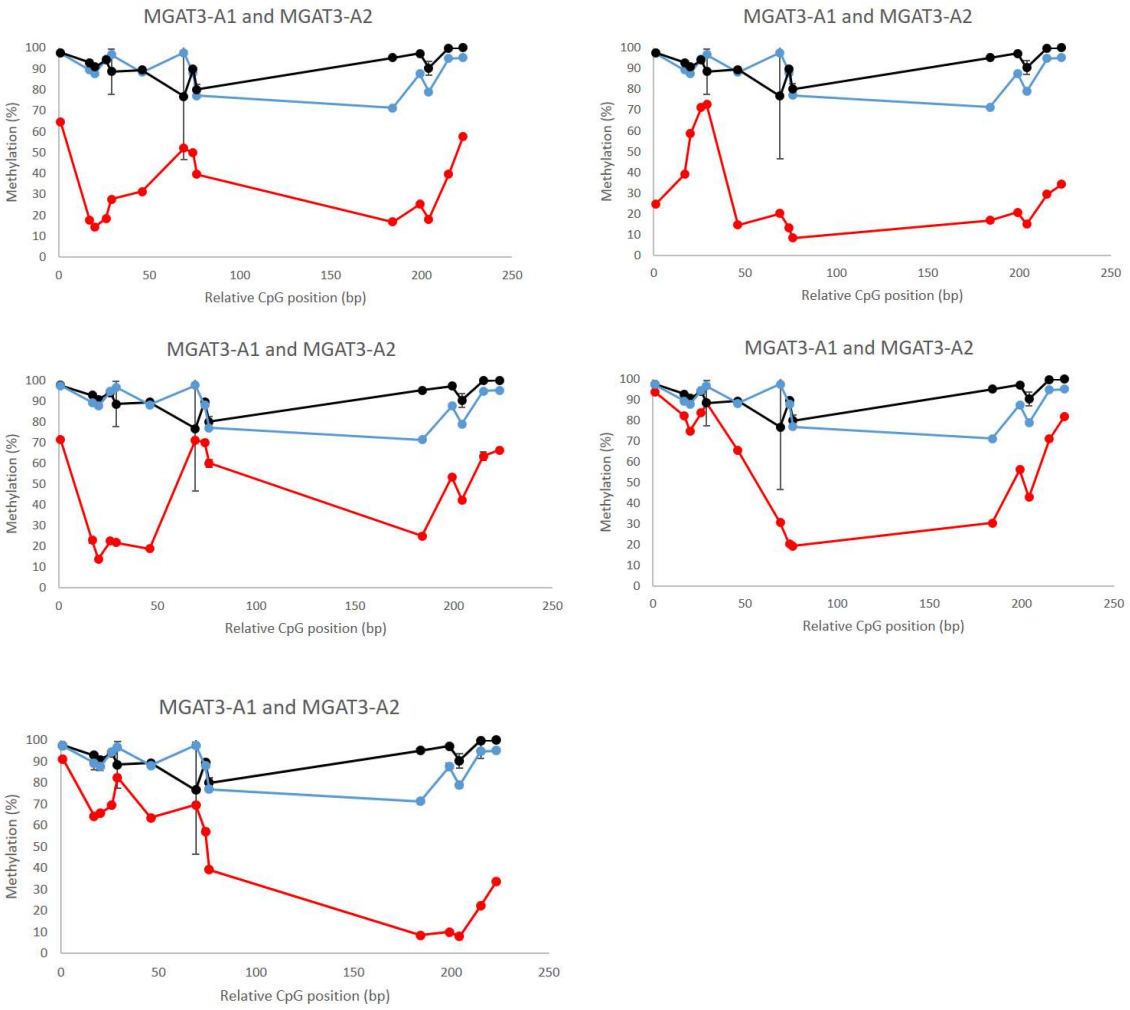


Figure 18. Targeted demethylation of the *MGAT3* promoter with TET1-dCas9 and 5 individual sgRNAs. Activity of TET1-dCas9 fusion protein, reflected on methylation level, guided by each individual sgRNA (MGAT3-sg01 – MGAT3-sg05) is shown as line connecting individual CpG sites (presented as dots). Same values for non-targeting control group (blue) and mock transfected cells (black) are shown in comparison to each sgRNA (red). Methylation status of each individual CpG site within the assays MGAT3-A1 and MGAT3-A2 was determined by bisulfite pyrosequencing. Relative CpG position and distance is plotted on X-axis (the assays MGAT3-A1 and MGAT3-A2 with total coverage of 111 bp

are shown together). Each dot represents an arithmetic mean of two replicas, with indicated standard deviation bars.

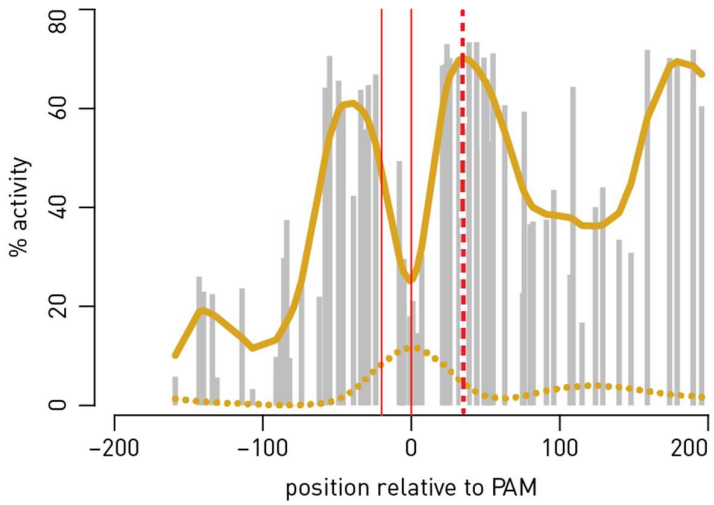


Figure 19. Summary profile of the TET1-dCas9 activity. Absolute methylation fraction decrease (compared to mock–transfected cells) is shown relative to the distance of each CpG site from the PAM sequence. Methylation fraction is indicated by orange line. Peak demethylation activity (red dotted line) was observed 30 bp downstream from PAM recognition sequence. Significant demethylation activity was also observed around 180-200 bp downstream of PAM sequence. Dotted orange line represents the absolute methylation fraction decrease for cells transfected with DED1-dCas9 in comparison with mock-transfected cells.

4.6 Time Course analysis of dCas9-TET1 demethylation

Methylation levels at CpG sties covered by the MGAT3-A1, MGAT3-A2 and LAMB1-A1 assays were analysed in HEK293 cells transfected with: (i) dCas9-TET1-MGAT3-sg03-PuroR, (ii) dCas9-DED1-MGAT3-sg03-PuroR, (iii) dCas9-TET1-LAMB1-sg01-PuroR, (iv) dCas9-DED1-LAMB1-sg01-PuroR, and (v) mock transfected cells. CpG methylation level was analysed at following time points: 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, 8th, 10th, 13th, 15th, 20th, 24th and 30th day post-transfection (transfection = day 0).

dCas9-TET1-MGAT3-sg03 targeting the *MGAT3* promoter resulted in methylation decrease for analysed CpG sites in the assays MGAT3-A1 and MGAT3-A2 in comparison with dCas9-DED1-MGAT3-sg03, dCas9-TET1-LAMB1-sg01 transfected cells and mock transfected cells. Methylation values for 9 CpG sites analysed within the MGAT3-A1 assay changed significantly between analysed time points (Figure 20). First significant methylation decrease (>5% methylation change at CpG sites 3-9) in comparison to mock transfected cells was observed on 3rd day post transfection, which correlates with complete 48 hours of puromycin selection and indicates a successfully enriched transfected cell population. Maximum demethylation activity was observed on 6th day after transfection for all analysed CpG sites (Figure 23). Highest methylation change of 29% was observed at third analysed CpG site of the MGAT3-A1 assay (6th day). Methylation changes dropped by 2-4% per CpG sites already on 7th day and slight increase in methylation was notable in every later time point. On 30th day, some CpG sites were still significantly demethylated in comparison to mock transfected cells; third CpG site had 19% lower methylation level, fifth CpG site had 9.6% and eight CpG site had 5.9% lower methylation level.

Methylation levels at 5 individual CpG sites analysed within the MGAT3-A2 assay changed significantly between analysed time points (Figure 21). First significant methylation decrease in comparison to mock transfected cells (>5%) was observed on the second day, still during puromycin selection, at third CpG site (7.9%). On the third day (immediately after complete puromycin selection) first three CpG sites all had >10% lower methylation levels. Maximum demethylation activity was observed on 6th day after transfection, with similar values observed for the seventh day. Highest methylation change of 54.3% at first analysed CpG site in the MGAT3-A2 assay was observed on the 7th day (Figure 23). Methylation is already slightly increased (around 2% per each CpG site) on the 8th day, and gradual increase in methylation is notable in every later time point. First and third CpG sites in the MGAT3-A2 assay remained significantly demethylated in comparison to mock transfected cells even on the 30th day (10.6% and 17.3% methylation changes).

dCas9-TET1-LAMB1-sg01 targeting the *LAMB1* promoter resulted in decrease of methylation level at 6 analysed CpG sites in the LAMB1-A1 assay in comparison with dCas9-DED1-LAMB1-sg01, dCas9-TET1-MGAT3-sg03 transfected cells and mock transfected cells. Methylation levels of individual analysed CpG sites significantly changed between

analysed time points (Figure 22). First significant methylation decrease (>5%) was observed on the second day with 5.74% methylation decrease at first CpG site and 6.94% at second CpG site in the LAMB1-A1 assay. Similar maximum methylation changes were observed on the 7th and 8th day, with greatest methylation change of 49.4% observed at first CpG site on the 8th day (Figure 23). Methylation level was already slightly increased (around 2% per each CpG site) on the 10th day, and gradually increased at every later time point. All 6 analysed CpG sites remained significantly demethylated (11.1% - 30.4%) in comparison with mock transfected cells even on the 30th day.

In this experiment, a different approach was used for assessment of non-specific, off-target construct activity. Instead of using dCas9-TET1-sgNT transfected cells as a control group, off-target activity was assessed by methylation analysis of the MGAT3-A2 assay for cells transfected with dCas9-TET1-LAMB1-sg01 (Figure 24) and by methylation analysis of the LAMB1-A1 fragment for cells transfected with dCas9-TET1-MGAT3-sg03 (Figure 25). Relatively high off-target activity was observed at certain CpG sites at certain time points.

For the MGAT3-A1 assay, first significant off-target demethylation change (>5%) was observed at third CpG site already on the 2nd day. On the 3rd day, first three CpG sites all had >7% lower methylation levels in comparison with mock transfected cells. Highest off-target demethylation of 26.7% was observed for first CpG site on the 7th day. First and third CpG sites remained significantly demethylated till the 24th day, with third CpG site having 5.2% lower methylation level in comparison with mock transfected cells even on the 30th day (Figure 24).

Within the LAMB1-A1 assay, first significant off-target methylation change (>5%) was observed the 3rd day at first, second, third and fifth CpG site. On fourth day, all analysed CpG sites showed >10% of methylation level decrease. Highest off-target demethylation was observed on 7th day (24.6% for second CpG site and 25.1% for fifth CpG site), with similar values observed on the 6th, 8th and 10th day. First four CpG sites remained significantly demethylated on 30th day (5.5%-11.2% methylation changes) (Figure 25).

CpG methylation in the MGAT3-A2 and LAMB1-A1 assays was analysed for cells transfected with dCas9-DED1 and targeting sgRNA expression plasmids in all 14 time points. No significant methylation changes were observed at any time point at any CpG site analysed

within the LAMB1-A1 assay. First and third CpG site in the MGAT3-A2 assay experienced significant demethylation at certain time points; first CpG site having 7.8% methylation change at the 6th day and remaining significantly demethylated till the 13th day (5.7%) and third CpG site having 5.3% methylation change at the 8th day and remaining demethylated till the 30th day (6.9 %) (Figure 26).

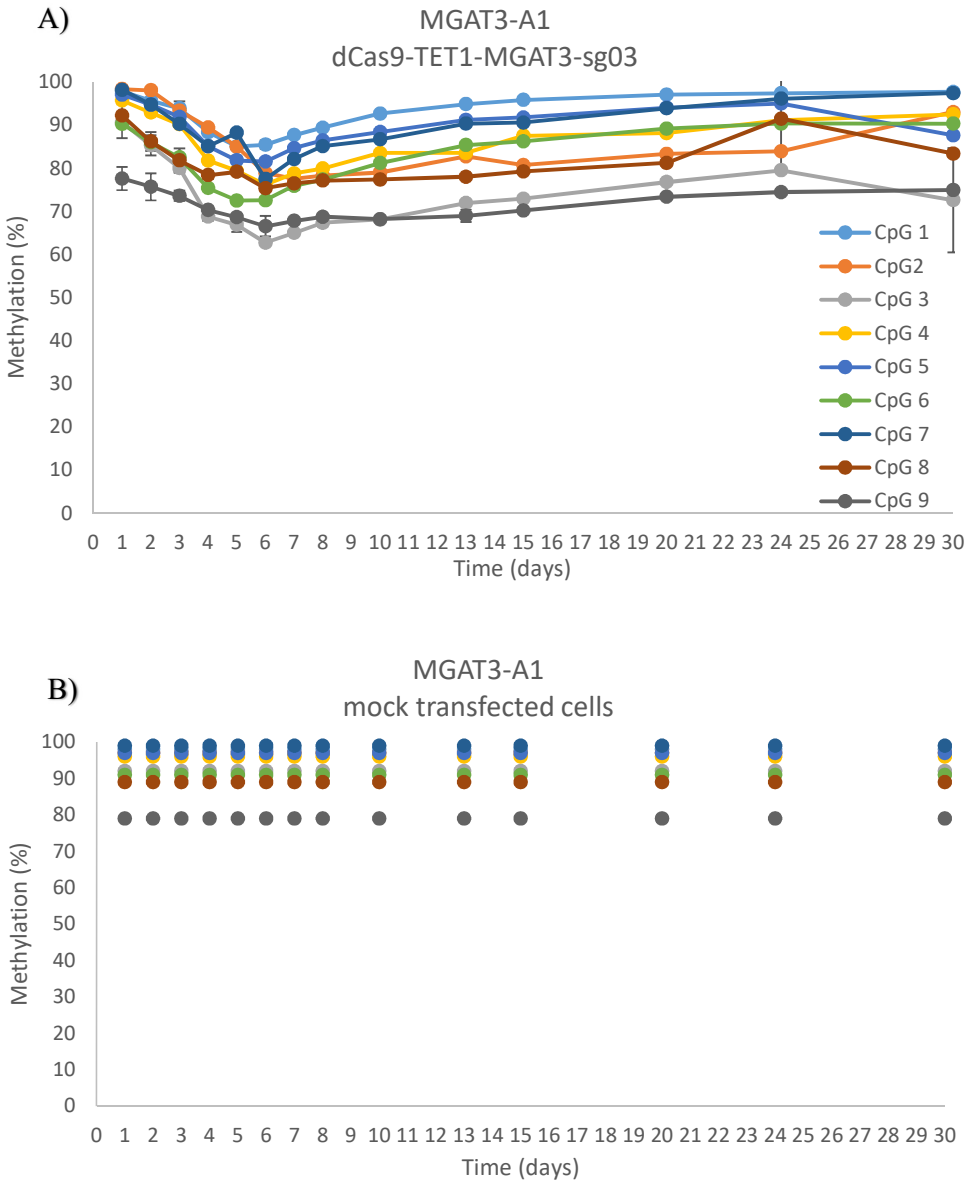


Figure 20. Time course evaluation of dCas9-TET1-MGAT3-sg03 induced demethylation effect on individual CpG sites within the MGAT3-A1 assay. A) Methylation level per analysed CpG site in the MGAT3-A1 assay was analysed at 14 time points between the 1st

and the 30th day after transfection with dCas9-TET1-MGAT3-sg03 expression plasmid. Methylation levels are presented as lines connecting individual time points (presented as dots). Each dot represents an arithmetic mean of methylation values of two replicas in a single time point, with indicated standard deviation bars. **B)** Methylation levels for each analysed CpG site in mock transfected cells are shown for comparison. Each dot represents an arithmetic mean of methylation values of two replicas in a single time point, with indicated standard deviation bars.

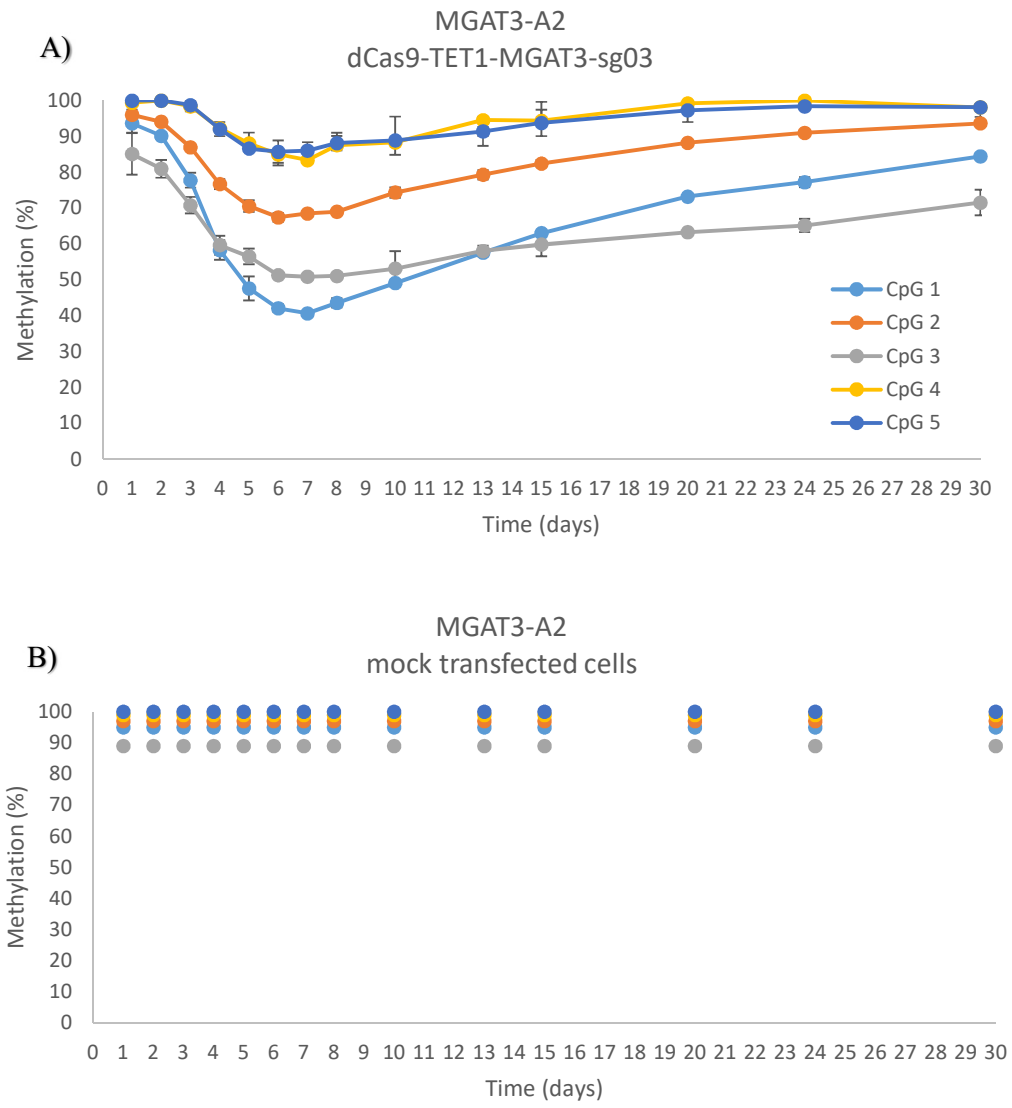


Figure 21. Time course evaluation of dCas9-TET1-MGAT3-sg03 induced demethylation effect on individual CpG sites within the MGAT3-A2 assay. A) Methylation level per

analysed CpG site in the MGAT3-A2 assay was analysed at 14 time points between the 1st and the 30th day after transfection with dCas9-TET1-MGAT3-sg03 expression plasmid. Methylation levels are presented as lines connecting individual time points (presented as dots). Each dot represents an arithmetic mean of methylation values of two replicas in a single time point, with indicated standard deviation bars. **B)** Methylation levels for each analysed CpG site in mock transfected cells are shown for comparison. Each dot represents an arithmetic mean of methylation values of two replicas in a single time point, with indicated standard deviation bars.

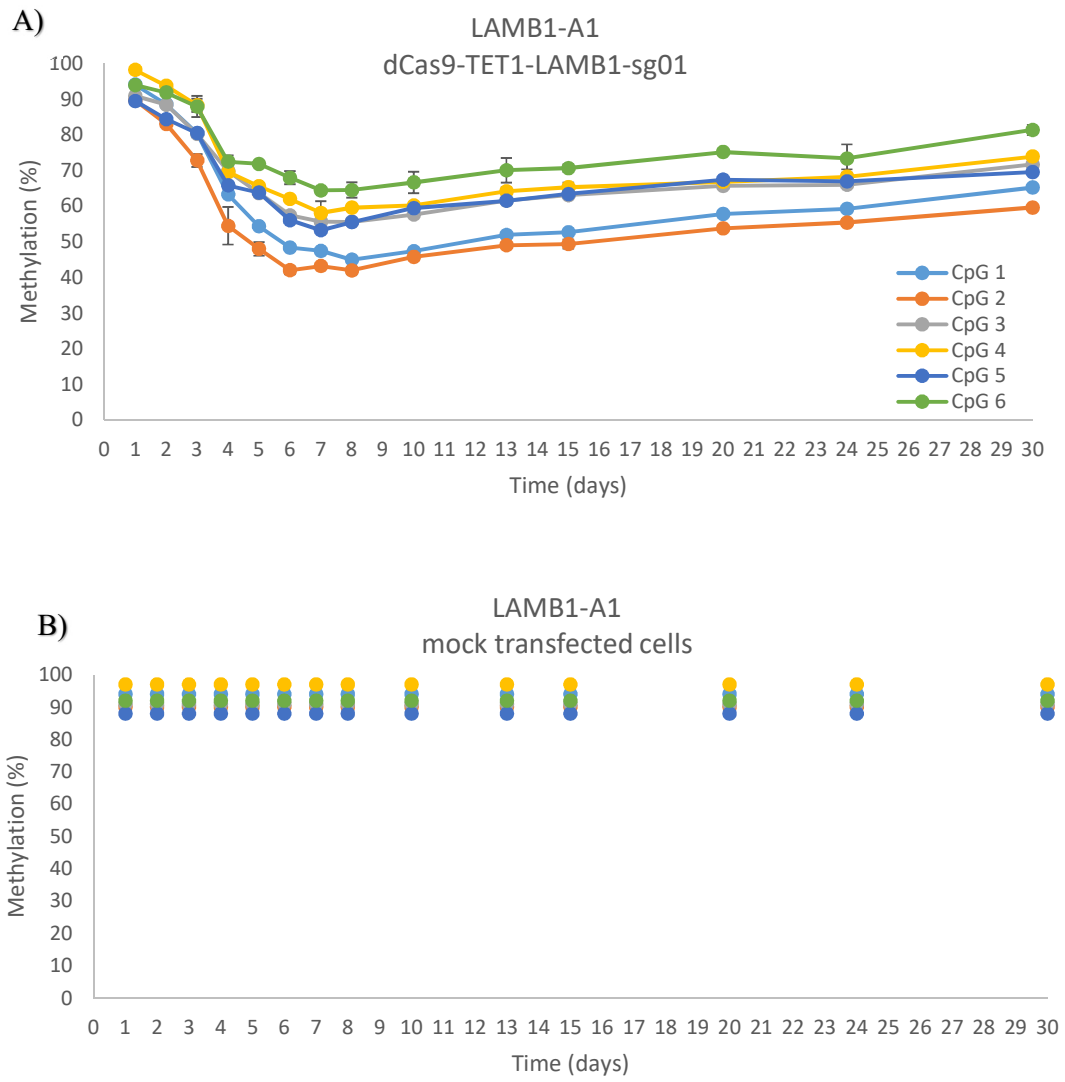


Figure 22. Time course evaluation of dCas9-TET1-LAMB1-sg01 induced demethylation effect on individual CpG sites within the LAMB1-A1 assay. A) Methylation level per

analysed CpG site in LAMB1-A1 assay was analysed at 14 time points between 1st and 30th day after transfection with dCas9-TET1-LAMB1-sg01 expression plasmid is shown. Methylation levels are presented as lines connecting individual time points (presented as dots). Each dot represents an arithmetic mean of methylation values of two replicas in a single time point, with indicated standard deviation bars. **B)** Methylation levels for each analysed CpG site in mock transfected cells are shown for comparison. Each dot represents an arithmetic mean of methylation values of two replicas in a single time point, with indicated standard deviation bars.

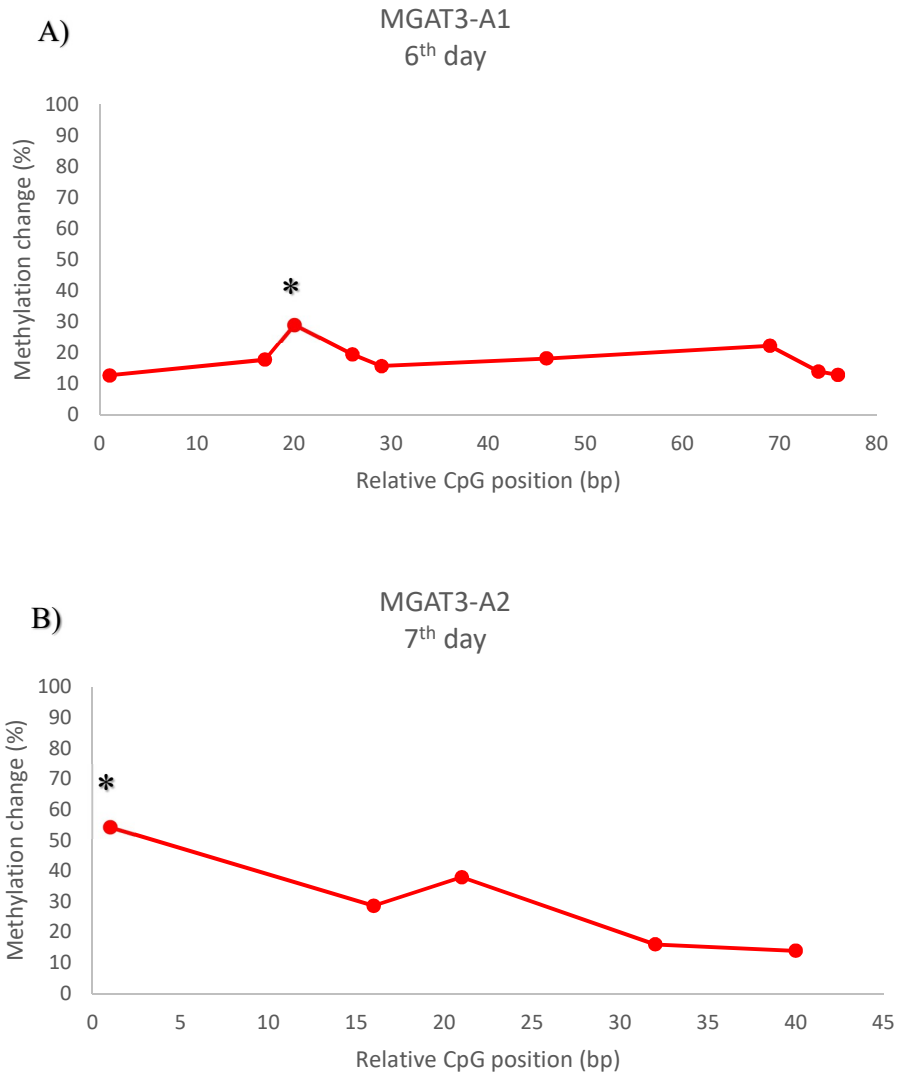


Figure 23. Figure is continued and described on the following page.

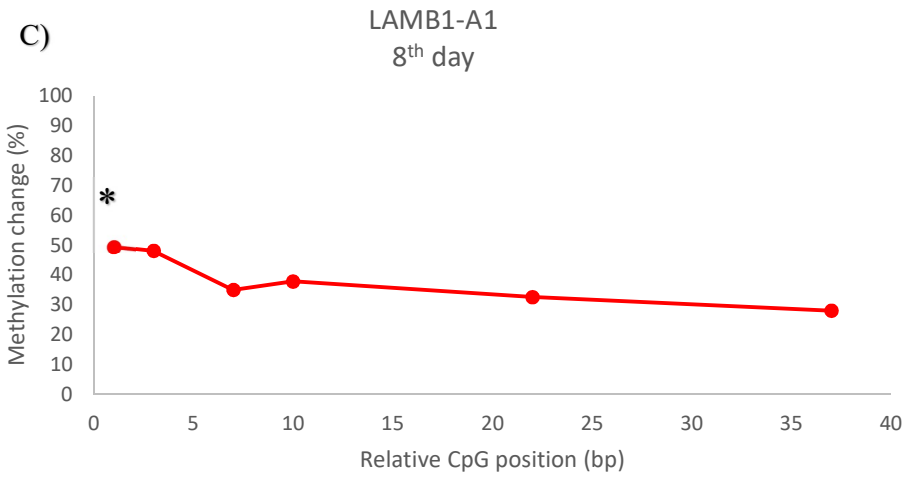


Figure 23. Methylation change per each CpG site in all analysed fragments (assays) at respective time points correlating with the highest observed demethylation activity. A) Methylation change per each CpG site within the MGAT3-A1 assay in comparison to mock transfected cells at sixth day after dCas9-TET1-MGAT3-sg03 transfection is shown. Highest methylation change of 29% at third CpG site is indicated (*). Bp distance between individual CpG sites is plotted on X-axis. **B)** Methylation change per each CpG site within the MGAT3-A2 assay in comparison to mock transfected cells at seventh day after dCas9-TET1-MGAT3-sg03 transfection is shown. Greatest observed methylation change of 54.3% at first CpG site is indicated (*). Bp distance between individual CpG sites is plotted on X-axis. **C)** Methylation change per each CpG site within the LAMB1-A1 assay in comparison to mock transfected cells at eighth day after dCas9-TET1-LAMB1-sg01 transfection is shown. Highest observed methylation change of 49.4% at first CpG site is indicated (*). Bp distance between individual CpG sites is plotted on X-axis.

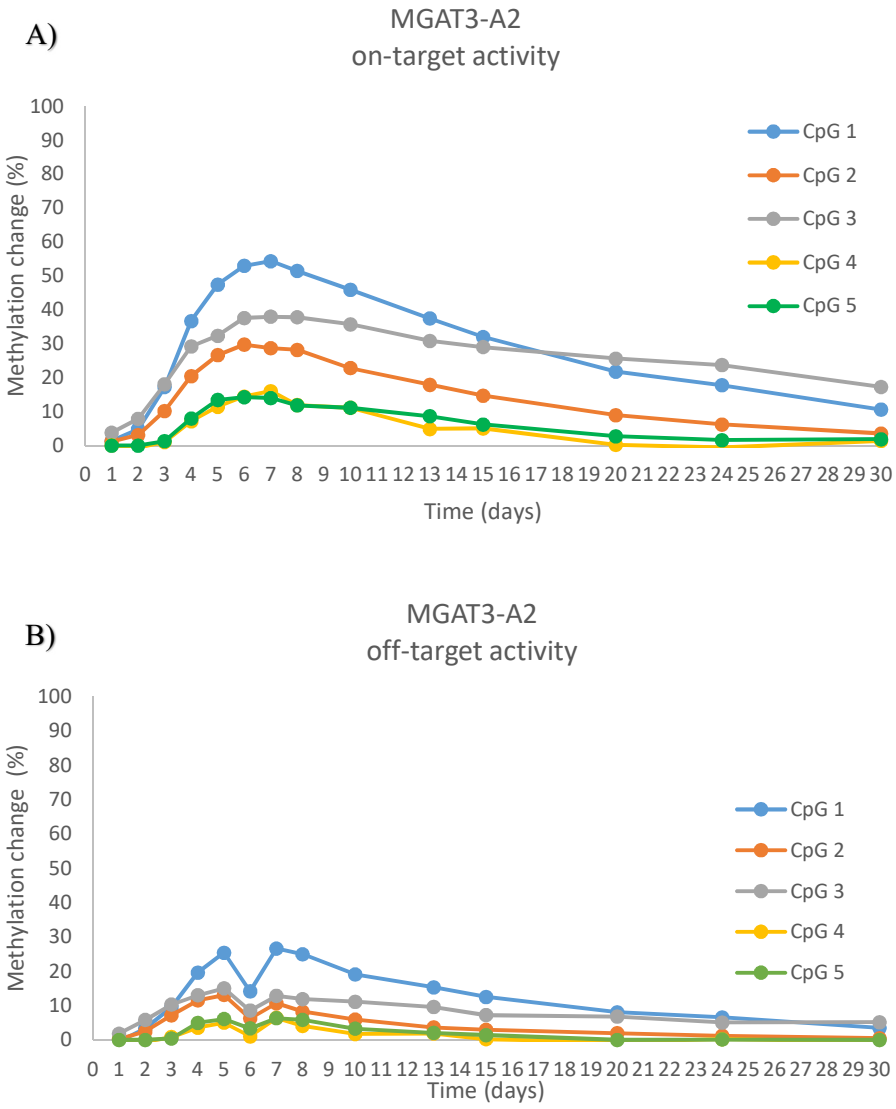


Figure 24. Comparison of on-target and off-target dCas9-TET1 demethylation activity at the MGAT3-A2 assay. A) On-target activity is represented as absolute methylation fraction decrease per CpG site in dCas9-TET1-MGAT3-sg03 transfected cells compared to mock transfected cells. First significant on-target methylation decrease of 7.9% at third CpG site was observed on second day. On the 3rd day, first three CpG sites exhibited >10% methylation change. Highest methylation change (54.3% at first CpG site) was observed on the 7th day after transfection, with similar values observed for sixth and eighth day. Gradual decrease in methylation level is observed in every later time point, with first and third CpG sites remaining significantly demethylated on the 30th day. **B)** Off-target activity is

represented as absolute methylation fraction decrease per CpG site in dCas9-TET1-LAMB1-sg01 transfected cells compared to mock transfected cells. First significant off-target demethylation change (>5%) was observed for third CpG site on second day. First three CpG sites exhibit >7% methylation change on the 3rd day. Highest off-target demethylation of 26.7% was observed for first CpG site on 7th day. Third and first CpG sites remained significantly demethylated until the 24th day, with third CpG remaining significantly demethylated on the 30th day.

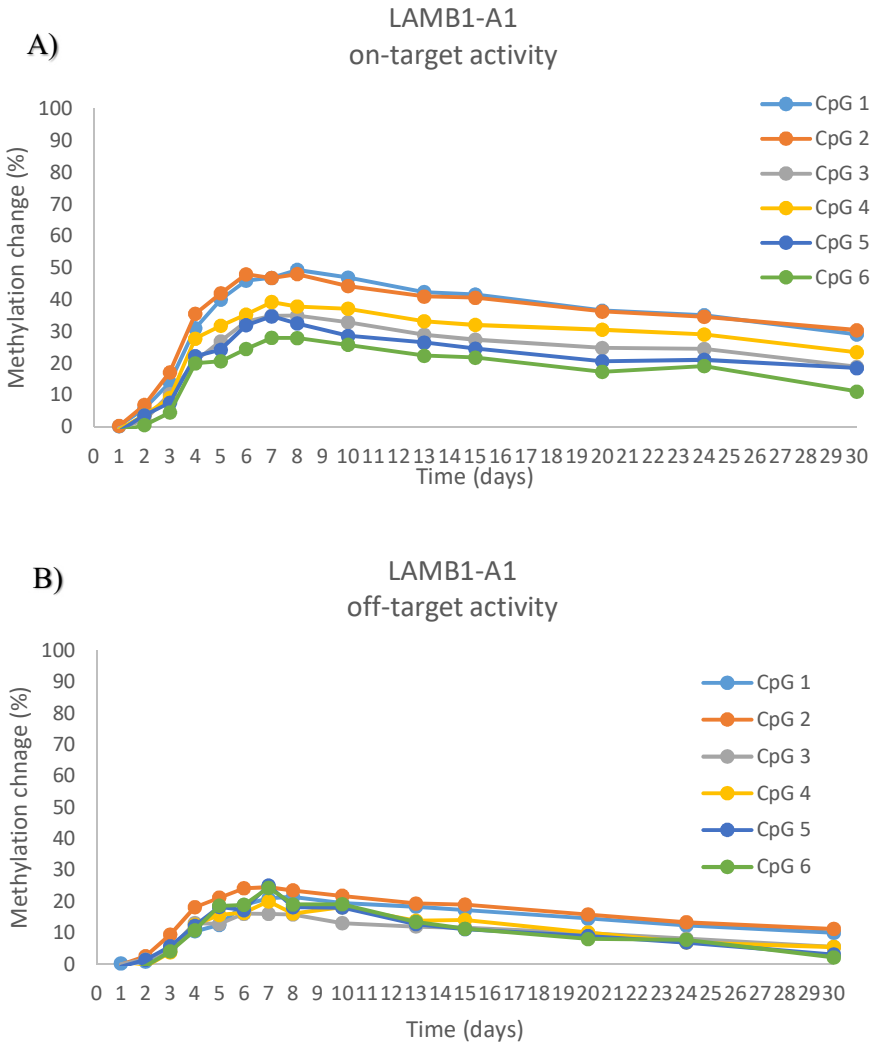


Figure 25. Comparison of on-target and off-target dCas9-TET1 demethylation activity observed for the LAMB1-A1 assay. A) On-target activity is represented as absolute methylation fraction decrease per CpG site in dCas9-TET1-LAMB1-sg01 transfected cells

compared to mock transfected cells. First significant on-target methylation decrease of 5.74% at first CpG site and 6.94% at second CpG site was observed on second day. Greatest methylation change (49.4% at first CpG site) was observed on 8th day after transfection. Gradual decrease in methylation is notable in every later time point. All 6 analysed CpG sites remained significantly demethylated (11.1% - 30.4%) in comparison with mock transfected cells even on the 30th day. **B)** Off-target activity is represented as absolute methylation fraction decrease per CpG site in dCas9-TET1-MGAT3-sg03 transfected cells compared to mock transfected cells. First significant off-target demethylation change (>5%) was observed on third day at first, second, third and fifth CpG site. All analysed CpG sites had >10% methylation change by fourth day. Greatest off-target demethylation was observed on 7th day (24.6% for second CpG site and 25.1% for fifth CpG site). First four CpG sites remained significantly demethylated on 30th day.

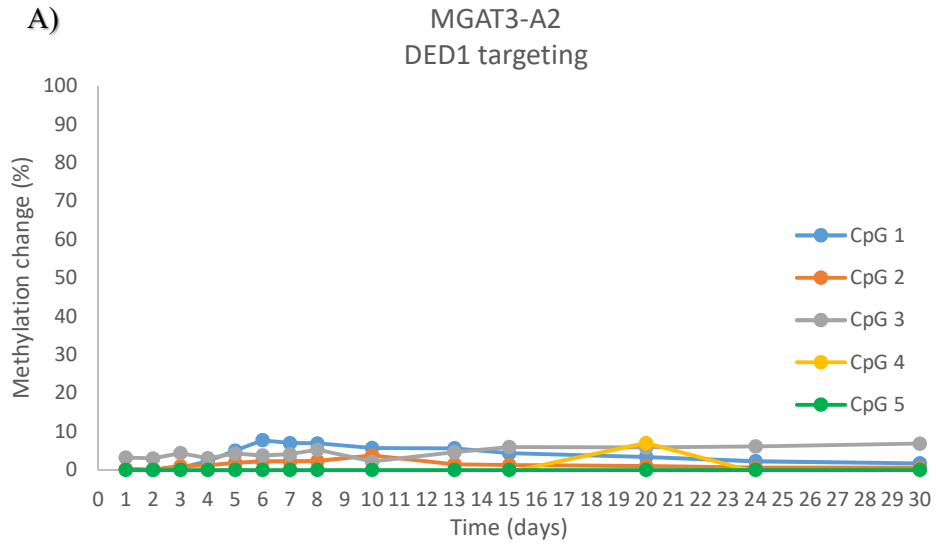


Figure 26. Figure is continued and described on the following page.

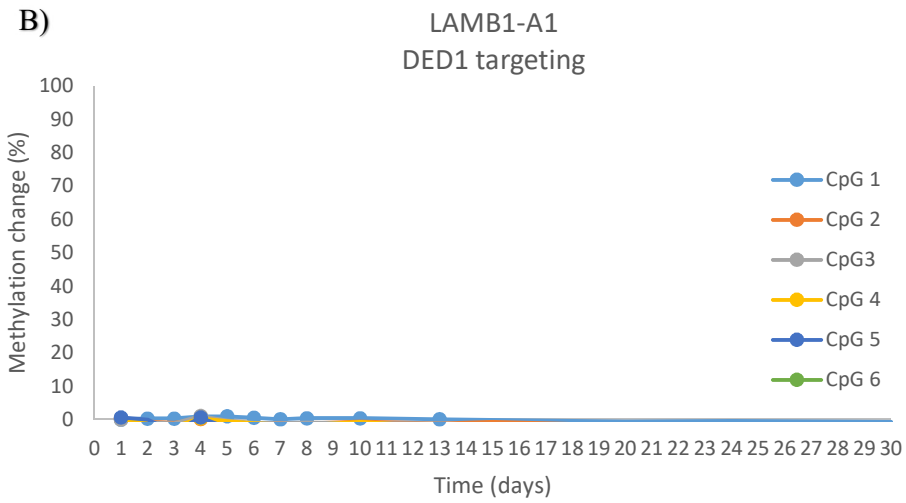


Figure 26. Methylation change per individual CpG sites in the MGAT3-A2 and LAMB1-A1 assays in cells transfected with dCas9-DED1 and sgRNA expression plasmids. Methylation change is represented as absolute methylation fraction decrease per CpG site in transfected cells compared to mock transfected cells. **A)** First and third CpG site of the MGAT3-A2 assay experienced significant demethylation at certain time points; first CpG site having 7.8% methylation change at the 6th day and remaining significantly demethylated till the 13th day (5.7%); third CpG site having 5.3% methylation change at the 8th day and remaining demethylated till 30th day (6.9%). **B)** No significant methylation changes were observed at any time point at any CpG site analysed within the LAMB1-A1 assay. Most of the methylation values obtained for dCas9-DED1-LAMB1 samples were either 1-2% higher or lower than those for mock transfected cells, thus those points were excluded from the representation.

4.7 Analysis of hydroxymethylation level

Methylation level at CpG sites within the MGAT3-A1 assay was analysed by bisulfite pyrosequencing for cells transfected with plasmids co–expressing dCas9-TET1-PuroR and sgRNA MGAT3-sg01, plasmids co–expressing dCas9-DED1-PuroR and sgRNA MGAT3-sg01, plasmids co–expressing dCas9-TET1-PuroR and non-targeting sgRNA and untransfected cells. One microgram of genomic DNA from each sample was selectively chemically oxidized to achieve 5-hydroxymethylcytosine (5hmC) to 5-formylcytosine (5fC) conversion which enables bisulfite conversion of 5fC to uracil. General denaturation and oxidation protocol was adapted from previously published work on hydroxymethylation level analysis, following instructions for genomic DNA manipulation (122). Oxidized DNA was bisulfite converted, amplified and methylation in the MGAT3-A1 assay was assayed as described above. Same procedure (bisulfite conversion, amplification and pyrosequencing) was done with the same amount of unoxidized isolated DNA for each sample and methylation levels of oxidized and unoxidized samples were subtracted to gain information about hydroxymethylation levels. Detection of any methylation difference after oxBS between samples of untransfected cells would indicate 5hmC deposition at this genomic region.

No significant methylation level changes were observed between oxidized and unoxidized DNA from dCas9-TET1 transfected cells, as well as between oxidized and unoxidized genomic samples from untransfected HEK293 cells (Figure 27). Unusual differences were observed in methylation levels for certain technical replicas (PCR products amplified from the same sample) between pyrosequencing runs on oxidized samples, which might imply inadequate sample purification between different steps of the oxBS procedure or that the oxidation procedure resulted in sample degradation in some extent.

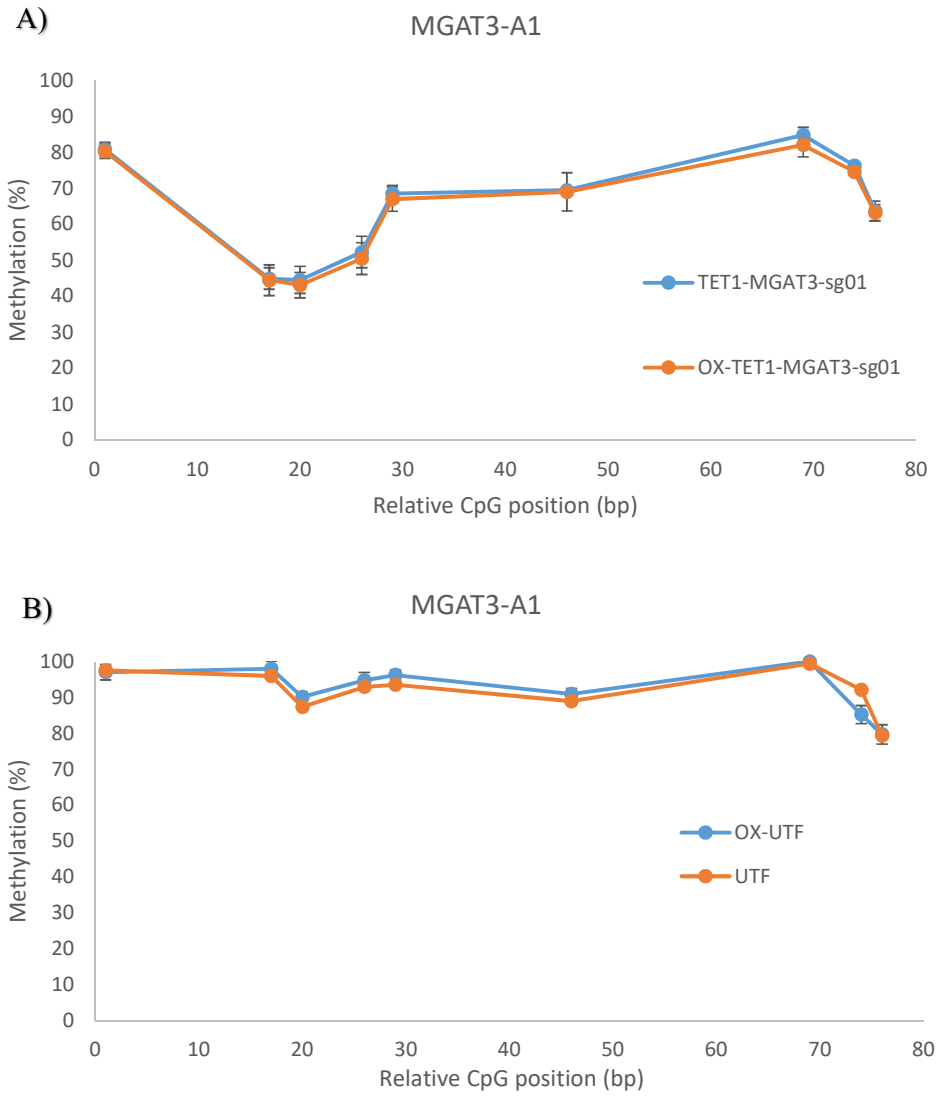


Figure 27. Methylation levels of oxidized and unoxidized samples determined by bisulfite pyrosequencing. A) DNA from cells transfected with dCas9-TET1-MGAT3-sg01 was analysed in its oxidized and unoxidized form. Methylation levels per each CpG site (dots) of MGAT3-A1 are shown as a line connecting individual sites. Each dot represents a mean of three replicas, each pyrosequenced two times, with bars indicating standard deviations. TET1-MGAT3-sg01 = DNA isolated from cells transfected with dCas9-TET1-MGAT3-sg01; OX-TET1-MGAT3-sg01 = oxidized DNA isolated from cells transfected with dCas9-TET1-MGAT3-sg01. **B)** Methylation levels per each CpG site of the MGAT3-A1 assay are shown for DNA from the untransfected cells analysed in its oxidized and unoxidized state with

bisulfite pyrosequencing. Each dot represents a mean of 3 replicas, each pyrosequenced two times, with bars indicating standard deviations. OX-UTF = oxidized DNA isolated from untransfected cells; UTF = DNA isolated from untransfected cells.

4.8 Targeting the *MGAT3* promoter with dCas9-TET1 and pooled sgRNAs

DNA and RNA were simultaneously isolated 8th day after transfection from the same cell pools. Biological groups were: i) HEK293 cells transfected with equimolar mixtures of five plasmids co-expressing dCas9-TET1 and a targeting sgRNA (MGAT3-sg01-sg05); ii) cells transfected with equimolar mixtures of five plasmids co-expressing dCas9-DED1 and a targeting sgRNA (MGAT3-sg01-sg05); iii) cells transfected with equal amount of plasmid co-expressing dCas9-TET1 and sgNT; iv) mock transfected cells. Isolated genomic DNAs were used for methylation analysis using the MGAT3-A2 assay and isolated total RNAs were reverse transcribed and used for *MGAT3* gene expression analysis.

Targeting of the *MGAT3* promoter with dCas9-TET1 and pooled sgRNAs resulted in methylation decrease at analysed CpG sites in the MGAT3-A2 assay in comparison with dCas9-DED1 and pooled sgRNAs transfected cells, dCas9-TET1-sgNT transfected cells and mock transfected cells (Figure 28). It is noticeable that targeting of *MGAT3* promoter with dCas9-TET1 and pooled sgRNAs resulted in higher demethylation activity in comparison with previous experiments where dCas9-TET1 was guided by any individual sgRNA. Highest methylation change was observed at first CpG site (77%).

MGAT3 mRNA levels were determined by qRT-PCR and normalized to endogenous control *HPRT1* using the comparative ddCt method (142). Changes in *MGAT3* gene expression within separate biological groups are expressed as fold change values in comparison to mock transfected cells (Figure 29). Transfection with dCas9-TET1 and pooled *MGAT3* targeting sgRNAs resulted in statistically significant increase (P=0.00004) of *MGAT3* mRNA of around 5-fold in comparison with mock transfected cells. Transfection with dCas9-DED1 and pooled *MGAT3* targeting sgRNAs also resulted in statistically significant increase (P=0.002) of *MGAT3* mRNA of around 3-fold in comparison with mock transfected cells, as well as transfection with dCas9-TET1-sgNT (P=0.004). Notably, there is a statistically

significant difference between dCas9-TET1 transfected cells and cells transfected with either dCas9-DED1 (P=0.012) or dCas9-TET1-sgNT (P=0.007).

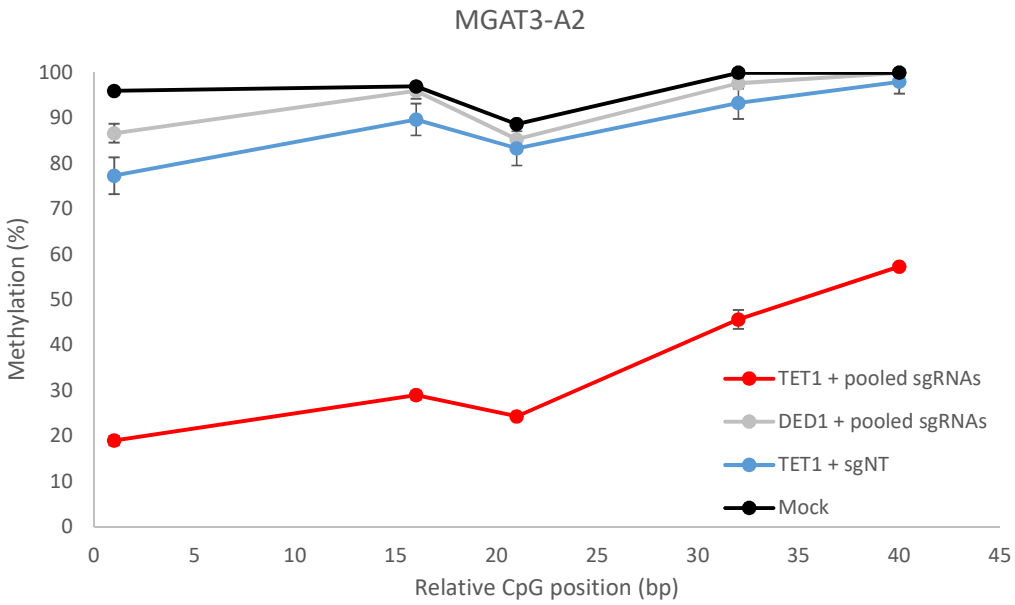


Figure 28. Targeted *MGAT3* promoter demethylation with dCas9-TET1 and pooled sgRNAs. Methylation levels at individual CpG sites within the MGAT3-A2 assay were determined by bisulfite pyrosequencing. Values for dCas9-TET1 and pooled sgRNAs, dCas9-DED1 and pooled sgRNAs, dCas9-TET1-sgNT and mock transfected cells are shown as lines connected by individual dots (CpG sites). Each dot represents an arithmetic mean of biological triplicates with indicated standard deviation. Bp distance between individual CpG sites is plotted on X-axis.

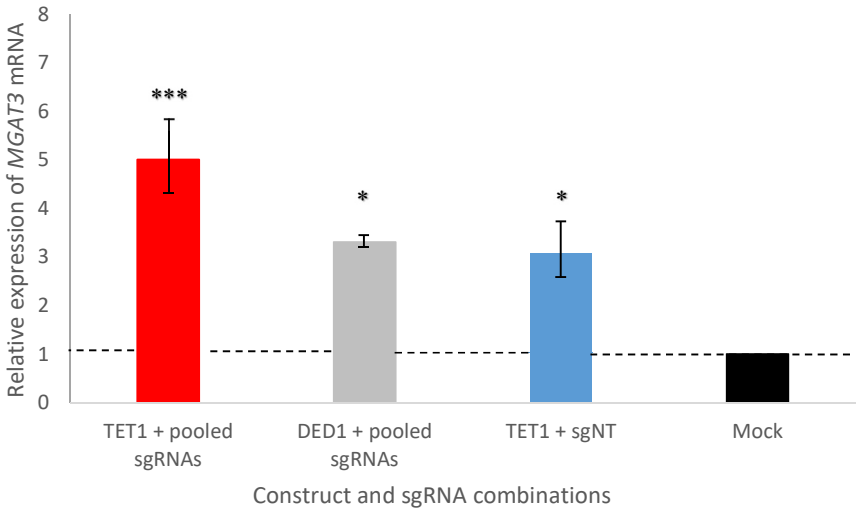


Figure 29. *MGAT3* gene expression changes after dCas9-TET1 targeted promoter demethylation. *MGAT3* mRNA expression levels were determined by qRT-PCR and normalized to endogenous control (*HPRT1*). Relative expression levels of *MGAT3* mRNA are shown as fold change values. Fold changes are shown relative to mock transfected cells (fold change=1). Mean values for biological triplicates are shown with indicated standard errors. Biological groups were compared among each other with ANOVA test followed by *post hoc* Tukey multiple comparisons test (***P ≤ 0.001; * P≤ 0.05).

4.9 Targeting the *MGAT3* gene with VPR-dCas9 and pooled sgRNAs

For VPR-dCas9 targeting of *MGAT3* gene, N-terminal fusion of VPR activation domain and dSaCas9 (pN-VPR-d(Sa)Cas9-PuroR) previously constructed in our lab was used with three different sgRNAs. Biological groups were: i) HEK293 cells transfected with VPR-dCas9 and each individual sgRNA (MGAT3-sg_INT-1f, MGAT3-sg_INT-2r and MGAT3-sg_INT-3f) expression plasmids; ii) HEK293 cells transfected with equimolar mixture of three plasmids co-expressing VPR-dCas9 and a targeting sgRNA (MGAT3-sg_INT-1f, MGAT3-sg_INT-2r and MGAT3-sg_INT-3f); iii) mock transfected cells. Total RNA from transfected cells was isolated 5 days after transfection, reverse transcribed and used for *MGAT3* gene expression analysis.

MGAT3 mRNA levels were determined by qRT-PCR and normalized to endogenous control *HMBS* using the comparative ddCt method (142). Changes in *MGAT3* gene expression within separate biological groups are expressed as fold change values in comparison to mock transfected cells (Figure 30). Transfection with VPR-dCas9 and pooled *MGAT3* targeting sgRNAs resulted in statistically significant increase (P=0.0003) of *MGAT3* mRNA of around 2-fold in comparison with mock transfected cells. Transfections with VPR-dCas9 and each individual *MGAT3* targeting sgRNA did not result in statistically significant increase in *MGAT3* expression. VPR-dCas9 and MGAT3-sg_INT-1f transfection resulted in statistically significant decrease (P=0.002) of MGAT3 expression, VPR-dCas9 and MGAT3-sg_INT-2r transfection resulted in non-significant (P=0.216) slight decrease in *MGAT3* expression and VPR-dCas9 and MGAT3-sg_INT-3f transfection resulted in non-significant (P=0.556) slight increase in *MGAT3* expression.

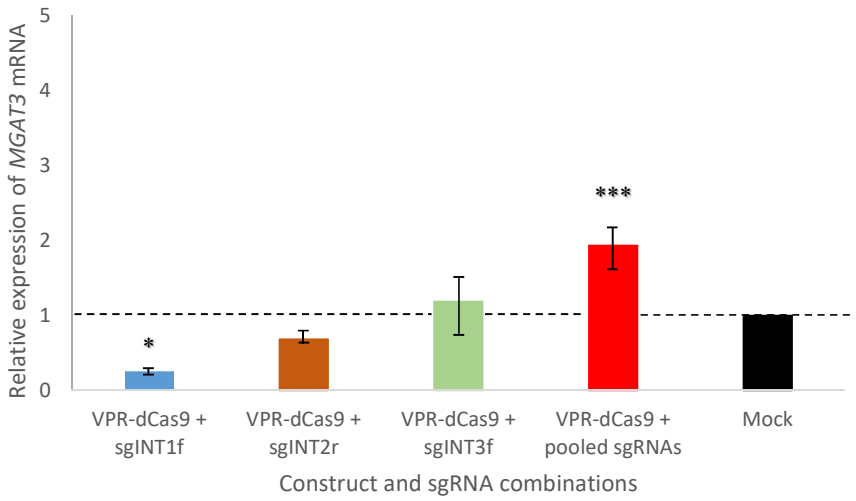


Figure 30. *MGAT3* gene expression changes following VPR-dCas9 targeting. *MGAT3* mRNA expression levels were determined by qRT-PCR and normalized to endogenous control (*HMBS*). Relative expression levels of *MGAT3* mRNA are shown as fold change values. Fold changes are shown relative to mock transfected cells (fold change=1). Mean values for biological triplicates are shown with indicated standard errors. Biological groups were compared among each other with ANOVA test followed by *post hoc* Tukey multiple comparisons test (***P ≤ 0.001; * P≤ 0.05).

5. Discussion

The aim of this doctoral thesis was development of flexible and easily reconfigurable molecular tools for targeted epigenetic manipulation and direct gene regulation. Special focus was given to development of CRISPR/dCas9-TET1 molecular tool for targeted DNA demethylation in mammalian genomes. Developed tools were validated on candidate loci *MGAT3* and *LAMB1*, genes associated with IgG glycosylation and inflammatory bowel diseases (18,23). I have validated the demethylation activity of both C-terminal and N-terminal fusions of TET1 catalytic domain (CD) and dCas9 from *Streptococcus pyogenes.* C-terminal dCas9-TET1 tool was generated by insertion of TET1 CD amplified from plasmid pJFA344C7 (map shown in Appendix 1) at the C-terminus of dCas9 via Gly_4Ser linker, and by addition of selection markers at the C-terminus of TET1 CD via short T2A peptide. N-terminal TET1-dCas9 tool was generated by insertion of TET1 CD amplified from constructed C-terminal dCas9-TET1 into a backbone for assembly of modular epitoolbox designed and created in our lab (Figure 14). VPR activation domain was inserted into the same backbone, resulting in VPR-dCas9 tool for CRISPRa. This tool was used for comparison of effects of the *MGAT3* promoter CpG demethylation on *MGAT3* expression and the effects of direct *MGAT3* gene activation with CRISPRa. TET1 CD sequence was amplified from TALE-TET1 fusion constructed and validated by Maeder et al (69). In their study, TALE fusions with both full length TET1 protein and TET1 CD were constructed and analysed, and those bearing minimal TET1 CD (Cys rich region and DSBH domain, as shown in Figure 7A) induced greatest CpG methylation changes. Different linkers were also tested in their study, with no significant variations in demethylation efficiency, so simple Gly_4Ser linker was applied. I used the same Gly_4Ser linker in construction of dCas9-TET1 and the fusion demonstrated strong demethylating activity. However, I did not test other linker lengths for dCas9-TET1 and TET1-dCas9 constructs, which may result in different profiles of construct catalytic activity. Nevertheless, the dCas9-TET1 tool constructed using Gly_4Ser linker showed strong demethylation activity varying from 18.8% to 68.3%, depending on a locus and/or a specific CpG site. The N-terminal TET1-dCas9 also induced demethylation, but with a remarkably stronger effect in comparison with dCas9-TET1. TET1-dCas9 exhibited strong demethylation activity varying from 40.8% to 87.1%, depending on a specific CpG site. Constructed tools are modular so that distinctive selection markers can

easily be cloned C-terminally of catalytic domain instead of PuroR and EGFP, such as hygromycin resistance gene or other fluorescent proteins. Due to the properties of the modular backbone used for insertion of TET1 and VPR domains in creation of TET1-dCas9 and VPR-dCas9 tools (Figure 14), these tools are highly modular and different sgRNAs, promoters, dCas9 orthologues, selection markers and terminators can be easily introduced in a single golden-gate reaction. Different effector domains can also be incorporated into the backbone in the same manner, so the system has an incredible potential for epigenetic editing.

Initial dose-response experiment showed greater demethylation levels in HEK293 cells transfected with more plasmid DNA, while non-specific demethylation activity remained similar for cells transfected with all tested plasmid amounts. An amount of 100 ng of plasmid DNA for transfections in 24-well plates (800 ng for 6-well plates) was chosen as optimal because it showed the greatest demethylation activity, best cell viability after transfection and minimal non-specific demethylation activity. The dCas9-TET1-EGFP construct was only used for assessment of transfection efficiency in initial experiments (data not shown), but it can be useful in transfections of sensitive cells which would not endure puromycin selection.

Several different sgRNAs were designed and cloned into dCas9-TET1 and TET1-dCas9 for *MGAT3* and *LAMB1* promoter targeting, as well as for VPR-dCas9 mediated *MGAT3* gene activation. There are specific guidelines for sgRNA design depending on the application of the CRISPR/Cas9 system (gene knockout, specific base editing, gene activation or repression and epigenetic modifications). For CRISPRa, which would be the use of VPR-dCas9, it is recommended to target a 100-200 bp window upstream of TSS (26). For targeted CpG demethylation with dCas9-TET1 and TET1-dCas9, it was necessary to pick sgRNAs providing a good coverage of the promoter fragments analysed by bisulfite pyrosequencing.

The *MGAT3* gene is encoded by two exons flanked by CGIs (Figure 9). Guide sequences for dCas9-TET1 targeting of *MGAT3* promoter were selected from all possible sequences that met obvious sgRNA design criteria (length of 20 bp, immediate downstream PAM sequence and confirmed high specificity with GT-Scan web-based tool (145) in a way to enable „targeting walking“ of promoter region that is covered by MGAT3-A1 and MGAT3-A2 pyrosequencing assays (Figure 9). Selected sgRNAs MGAT3-sg01 – MGAT3-sg05 were chosen to be all in close proximity of analysed CpG sites, but distant enough to cover a 190 bp region quite densely. MGAT3-sg07 and MGAT3-sg08 were located 260 bp

and 320 bp from last analysed CpG site in MGAT3-A2, and distance between MGAT3-sg01 and MGAT3-sg08 respective binding sites is 550 bp. Such sgRNA design enabled precise profiling of dCas9-TET1 activity and the same sgRNAs were used for profiling of N-terminal TET1-dCas9 fusion.

When targeting a 100 bp window upstream of TSS for CRISPRa (using VPR-dCas9), there were not so many potential sgRNAs to choose from. It was previously described for TALE-based activators that multiple targeting of endogenous genes results in stronger, synergistic gene activation (67). Same principle applies to CRISPR/Cas9 gene activation, most efficient activation is observed when combinations of sgRNAs binding in close proximity within the promoter region, thus covering a larger DNA region, are simultaneously expressed in the same cell (146,147). Same approach of multiple targeting was adopted for this study. Considering that SaCas9 has a PAM sequence of 5'-NNGRRT-3', which is less frequent than SpCas9 PAM (5'-NGG-3'), sgRNA design was even more difficult. Several sgRNAs were designed and cloned into the VPR-dCas9 tool for targeted *MGAT3* gene activation. First sgRNA picks were all in a 100 bp window from potential TSSs predicted around first CGI and the beginning of the first exon, but none of them showed any effect on *MGAT3* gene expression. So, *MGAT3* gene sequence was further screened for predicted regulatory elements, like promoters predicted by FANTOM database. A study by Radzisheuskaya and collaborators recently demonstrated that the FANTOM database provides the most reliable source of TSSs annotations (148). Several potential TSSs were predicted inside the intron of *MGAT3* gene, so several sgRNAs targeting separate intronic regions were designed and tested. Three sgRNAs finally selected and presented in this study gave a modest *MGAT3* gene activation effect of ≈2 fold, but tested best from all considered options. They all bind within the *MGAT3* intron, around 280 bp from the beginning of second exon and 450 bp upstream of the second CGI (Figure 11). Further screening of *MGAT3* gene for VPR-dCas9 induced gene activation would be needed. Interestingly, when tested individually, MGAT3-sg_INT-1f and MGAT3-sg_INT-2r resulted in a decrease of *MGAT3* expression, but when used together with MGAT3-sg_INT-3f (which resulted in a slight *MGAT3* activation when tested individually) gave a stronger, synergistic activation effect (Figure 28.)

Each sgRNA used for dCas9-TET1 targeting of *MGAT3* promoter induced variable demethylation of specific analysed CpG sites, depending on its binding site (Figure 16). All

five of the proximal sgRNAs targeting the *MGAT3* promoter induced a ≈50% decrease in methylation at certain CpG sites, and maximal demethylation effect of 68% for dCas9-TET1 targeting with an individual sgRNA was observed at the first CpG site of MGAT3-A2 for MGAT3-sg05. Much stronger synergistic demethylation was observed for combined targeting of *MGAT3* promoter with all five targeting sgRNAs, with maximal effect of 77% observed at first CpG of MGAT3-A2, but all of the analysed CpG sites were significantly more demethylated than in any case of individual sgRNA targeting (Figure 26).

N-terminal fusion of TET1 and dCas9 induced stronger demethylation effects for each of the tested sgRNAs (MGAT3-sg01 – MGAT3-sg05) in comparison with dCas9-TET1, with a similar activity profile. Average peak demethylation effect for each sgRNA targeting with TET1-dCas9 was around 76% for CpG sites 30 bp downstream from PAM, which is comparable to 77% demethylation effect observed with combined targeting of *MGAT3* promoter with all five targeting sgRNAs and dCas9-TET1. TET1-dCas9 also induced stronger demethylation activity upstream from PAM sequence, up to 76.9% for the CpG site 76 bp upstream from PAM of MGAT3-sg03, while C-terminal fusions also had considerable upstream demethylation activity but with maximal effect of 22.7% demethylation for the CpG site 57 bp upstream from PAM of MGAT3-sg03.

Summary activity profiles of dCas9-TET1 and TET1-dCas9 fusions were quite similar with TET1-dCas9 exhibiting approximately 20-40% stronger demethylation for the same CpG sites downstream from respective PAM recognition sequences and 20-54% stronger effects for the same CpG sites upstream from respective PAM sequences. Peak demethylation activity seemed to occur a bit further away from PAM for TET1-dCas9, both downstream (4-5 bp) and upstream (2-19 bp) in comparison with dCas9-TET1. N-terminal fusion also exhibited 5.5% stronger off-target effects for the same CpG sites than C-terminal dCas9-TET1 fusion. It would be interesting to analyse the activity of TET1-dCas9 and pooled sgRNAs (MGAT3-sg01 – MGAT3-sg05) and check for synergistic demethylation effect and *MGAT3* expression changes.

Another study demonstrated a successful fusion of the mouse Tet1 CD and dCas9 protein which was additionally empowered by sgRNA2.0-guidance strategy and MS2-Tet1-CD construct (149). Konermann and collaborators first introduced the sgRNA2.0 concept for improved CRISPRa. By analysing the previously determined crystal structure of the dSpCa9

in complex with sgRNA and targeted DNA (150) they observed and further demonstrated that some sgRNA loop structures do not affect dSpCas9 functions and could tolerate the addition of RNA hairpin aptamers which could facilitate the recruitment of additional effector domains to the Cas9 complex. They added two copies of bacteriophage MS2 RNA elements to two sgRNA loop structures. MS2 bacteriophage coat proteins (MCPs) can dimerize and bind specific RNA hairpin aptamers, so effector domains fused to MCPs can bind dCas9-sgRNA2.0 complexes for more efficient activation (151). Xu and collaborators demonstrated that sgRNA2.0, dCas9-Tet1-CD and MS2-Tet1-CD (Tet1 CD fused to MCP) targeting can induce specific demethylation at promoter regions of target genes and achieve substantial activation of gene expression (149). For sgRNA2.0 and dCas9-Tet1-CD and MS2-Tet1-CD targeting with simultaneously expressed multiple sgRNAs no synergistic demethylation activity was observed, in contrast with previous observations for CRISPRa tools (146) and in contrast with results obtained in this study for both dCas9-TET1 and VPR-dCas9 targeting. The authors have attributed this result to mechanistic differences between demethylation events catalysed by Tet1 CD and recruitment of transcription complexes by activators such as VP64. In my work, I did not observe such difference as both dCas9-TET1 and VPR-dCas9 experienced synergistic effects when simultaneously targeted with multiple sgRNAs. But then again, intentions with creating sgRNA2.0 systems were to achieve robust up-regulation with a single sgRNA. In some cases in the initial study, targeting with a single sgRNA2.0 and dCas9-VP64 paired with MS2-p65-HSF1 (activation domain from human heat-shock factor 1 fused to MCP) demonstrated more robust activation activity than targeting with dCas9-VP64 combined with a pool of eight classic sgRNAs (151).

The peak demethylation activity of the dCas9-TET1 tool constructed in this work was observed 26 bp downstream from PAM recognition sequence and significant demethylation activity was also observed around 180-200 bp downstream of PAM (Figure 17). Maeder and collaborators observed a quite similar profile for their TALE-TET1 fusions, with greatest decrease in methylation within 30 bp of end of the TALE target binding site and significant CpG demethylation around 150-200 bp away from the target site. They attribute this effect to TET1 CD accessing regions of open chromatin located at least a nucleosome away (69), which seems plausible and probably explains the similar effect observed with dCas9-TET1. A

slight peak in demethylation is noticeable approximately 25 bp upstream of PAM for dCas9-TET1, so this construct demethylates on both sides from its binding site (Figure 17).

Methylation levels of individual CpG sites between cells transfected with dCas9-TET1-sgNT did not significantly differ from those of mock transfected cells except for the first CpG site within the MGAT3-A2 assay, which was 18.5% demethylated in dCas9-TET1-sgNT transfected cells. Two sgRNAs, located more than 230 bp downstream of last analysed CpG site in MGAT3-A2 assay, which had no observable methylation changes for any of the analysed CpG sites, showed a methylation decrease for the first CpG site of MGAT3-A2 (20.8% for MGAT3-sg07 targeting and 13.2% for MGAT3-sg08 targeting). All of the other 5 sgRNAs used for dCas9-TET1 targeting of the *MGAT3* promoter experienced peak activity (36%- 68.3% methylation change) for MGAT3-A2 assay at first analysed CpG site. When taken together, these results suggest that this CpG site might be in a most accessible chromatin conformation for TET1 CD to act. It would be interesting to develop a pyrosequencing assay which would encompass 4 CpG sites located in a 106 bp region between the last analysed CpG site of MGAT3-A1 assay and first CpG site of MGAT3-A2 and assay the methylation level. Sometimes differential methylation of as little as four CpG sites can play a major role in gene expression regulation (152–154) so targeting and methylation analysis of these sites in particular might provide some interesting findings.

CpG methylation within promoter regions correlates with gene silencing, and unmethylated CpG dinucleotides within those regions usually indicate sites of active transcription (155). In this work, I have demonstrated that targeted CpG demethylation of specific CpG sites within *MGAT3* promoter resulted in ≈5-fold increase in its transcriptional expression (Figure 27). For this experiment, RNA from transfected cells was isolated 8 days after transfection. This time point was chosen because the previous time course experiment of dCas9-TET1 activity had shown the peak demethylation activity on either 6^{th}, 7^{th} or 8^{th} day after transfection. It might be informative to analyse the *MGAT3* gene expression in the earlier and later time points as well as to perform a complete expression analysis profile. Interestingly, the *MGAT3* gene expression was also increased by ≈3.3-fold in cells transfected with dCas9-DED1 targeted with the same pool of sgRNAs and by ≈3-fold in cells transfected with dCas9-TET1-sgNT. These observations might indicate that a decrease in methylation at first CpG site of MGAT3-A2 assay (9.3% for dCas9-DED1 targeted cells and 18.7% for

dCas9-TET1-sgNT transfected cells) has an impact on the *MGAT3* gene regulation. Effects observed for dCas9-DED1 targeting could not be explained undoubtedly, as one might expect a slight decrease in gene expression due to CRISPR interference in such case (156). The *MGAT3* gene is normally expressed in HEK293 cells at very low levels so I would expect such effect would not be noticeable. Increase in *MGAT3* expression after dCas9-DED1 targeting might be explained by recruitment of other activators which could normally interact with TET1 CD after its binding to DNA. A study by Jin and collaborators demonstrated that overexpression of both TET1 CD and its catalytically inactive variant changed gene expression profiles in a similar manner, even though only TET1 CD overexpression resulted in massive DNA demethylation. These findings suggest that even the catalytically inactive CD influences gene expression by its binding in a demethylation-independent way (136).

Both dCas9-TET1 and VPR-dCas9 approaches resulted in the *MGAT3* gene activation. It should be noted that RNA from VPR-dCas9 transfected cells was isolated on the 5th day after transfection and from the dCas9-TET1 transfected cells on the 8th day after transfection (when peak demethylation was expected). An additional time-course experiment would be needed for analysis of the *MGAT3* gene expression for both approaches, as such time-course analysis might provide information on which minimal methylation level changes at this locus are sufficient to alter gene expression. It would also be interesting to analyse methylation levels of the *MGAT3* promoter from cells transfected with VPR-dCas9 to see what effects this CRISPRa might have on promoter demethylation by endogenous proteins.

The dCas9-TET1 has all of the previously discussed advantages CRISPR/Cas9 system offers in comparison with ZF- and TALE-based tools, such as simplicity of programming and construction. For example, previously mentioned TALE-TET1 fusions of comparable activity (69) were also designed to bind 20 bp target sequences (like sgRNAs designed for dCas9-TET1 and TET1-dCas9), since this is the length considered long enough to provide specificity in human genome. However, construction of such sequence-specific proteins is much more laborious and time-consuming in comparison with simple sgRNA selection and cloning into once constructed dCas9-TET1 expression plasmid. Also, *in vitro* transcribed sgRNAs can be used (rather than sgRNAs expressed from plasmids) – an approach which omits cloning from Cas9 targeting (157). One of the disadvantages of CRISPR/dCas9 system is the relatively large size of the expression vectors (e.g. dCas9-TET1-PuroR is 11 373 bp large and dCas9-

TET1-EGFP is 11 484 bp large) which is a limiting factor for cell transfections. Although plasmid features of dCas9-TET1 tools, such as promoters and linkers, as well as a minimal catalytic domain of TET1, are quite compact and truncated to maximum, there are still size-limiting elements which cannot be omitted or truncated. For example, dSpCas9 itself is 4101 bp or 1367 aa large. Usage of T2A self-cleaving peptide is an elegant way of eliminating the need for two separate promoters and a nice vector minimizing strategy. However, an expression vector still contains a 4101 bp large dSpCas9 sequence, which make a construct very large when fused to a catalytic domain, even the smallest one such as that of TET1 which totals 2163 bp (seen in Figures 13 and 14). Size of the (d)SpCas9 is also an important limiting concern in therapeutic delivery due to packaging capacity of Adeno-associated viruses (AAV). Smaller orthologues of Cas9 protein can be used to circumvent this issue, for example Cas9 from *Neisseria meningitides* (NmCas9) which is 1082 aa large, or *Staphylococcus aureus* Cas9 which is 1053 aa large (59,158). However, these smaller Cas9 orthologues require more complicated PAM recognition sequences - NmCas9 recognizes a 5'-NNNNGATT-3' PAM and SaCas9 a 5'-NNGRRT-3' PAM, sequences which are harder to find in the genome than 5'-NGG-3'. Therefore, this factor limits the targeting capability of tools which use Cas9-orthologs. On the other hand, the Cas9 orthologues which recognize more complex and rare PAM sequences might provide reduced off-target effects. PAM sequence for NmCas9 occurs once every 128 bases in a random sequence, and it has been demonstrated that NmCas9 exhibited lower off-target effects on endogenous loci than SpCas9, while achieving similar on-target activities (158).

In this study, HEK293 cells were only transiently transfected with dCas9-TET1 and targeting sgRNA expression plasmids. Cells were selected with puromycin for 48 h to achieve a relatively homogenous population of cells successfully expressing dCas9-TET1, but in such transfections expression plasmids are lost from the cell pool with subsequent cell divisions. This was observed in a time-course experiment with a gradual increase in methylation from 10^{th} till 30^{th} day after transfection. It would also be interesting to correlate the presence of the plasmid DNA in cell populations over the time-course period with the level of the methylation changes, which I have not analysed. Plasmid quantification during a 30 day period was done with a previous dCas9-DNMT3A tool developed in our group (17) and I would expect for dCas9-TET1 to have a similar retention period in cell pools due to similar plasmid features.

Peak demethylation activity was observed on the 6th or 7th day in case of the *MGAT3* gene and on the 8th day (but with similar values for 7th day) in case of the *LAMB1* gene targeting. First significant demethylation in some of the CpG sites in both promoters was observed on 2nd and 3rd day after transfection, correlating with 24 h or 48 h of puromycin selection, when dCas9-TET1 expressing cells should be prevalent in cell pools. Slight decrease in demethylation was observed on 10th day, with a gradual decrease in every later time point for both analysed genes. These findings suggest similar dCas9-TET1 demethylation pattern in different loci. However, for precise profiling more endogenous loci should be targeted with this molecular tool and analysed for methylation level in additional time-course studies. Interestingly, for analysed *MGAT3* fragments several CpG sites remained significantly demethylated in comparison to mock transfected cells even on the 30th day after transfection (3rd, 5th and 8th CpG of MGAT3-A1 and 1st and 3rd CpG site of MGAT3-A2). Also, all of the analysed CpG site within LAMB1-A1 assay had 11.1-30.4% lower methylation levels than mock transfected cells on 30th day. Maeder and collaborators performed a time-course evaluation of the effect of TALE-TET1 tool on CpG methylation level in *HBB* gene promoter region on 4th, 7th, 14th and 30th day after transfection. They observed a peak demethylation activity on the 4th day, with a gradual increase in methylation till 30th day (69). This can also be attributed to the loss of expression plasmid through cell divisions, but all CpG sites they analysed reverted to methylation levels of control cells by 30th. Kungulovski and collaborators analysed the stability of individual DNA and histone methylation changes induced by targeting with DNA and histone methyltransferases and found that methylation, chromatin conformations and gene expression levels returned to their original state after methyltransferases were lost from cell pools. They suggested that several distinctive epigenetic marks should be introduced to the target site for epigenetic reprogramming to be stable or epigenome editing tools should be expressed in cells for a longer period of time (159). Rivenbark and collaborators generated stable cell lines containing the gene of the targeted methyltransferase and induced DNA methylation and repression of *Maspin* gene stable over multiple generations (64). Xu and collaborators noted much higher gene activation rates (10- to 20-fold) with stable expression of the sgRNA2.0, dCas9-Tet1-CD and MS2-Tet1-CD system compared to transiently transfected system (2-fold) (149).

Two different sequencing-based approaches have been developed for 5hmC detection. Yu and collaborators described a Tet-assisted bisulfite sequencing (TAB-seq) method for detection of 5hmC at single-base resolution. This method relies on β-glucosyltransferase-mediated protection of 5hmC by glucosylation and recombinant mouse Tet1-mediated oxidation of 5mC to 5caC, followed by bisulfite conversion and PCR amplification. In this approach, 5hmC reads as C, while all unmethylated cytosines and 5mC (turned to 5caC) are converted and read as T (160). Alternative approach is oxidative bisulfite sequencing (oxBS-seq), in which 5hmC is selectively oxidized to 5fC by chemical treatment with an oxidant like $KRuO_4$, followed by bisulfite conversion and amplification. In this approach, 5mC is read as C, while all unmethylated cytosines and 5mhC (turned to 5fC) are bisulfite converted and read as T (122). In this study the oxBS-seq approach was adopted as it is cheaper than TAB-seq for it does not require recombinant Tet1 enzyme. I have followed the instructions for genomic DNA oxidation with $KRuO_4$ as detailed in the work of Booth et al. (122). Authors claim that the oxBS-seq workflow can be used to modify DNA for analyses by Sanger sequencing, pyrosequencing, high-throughput sequencing and methylation arrays such as Illumina 450K arrays. Hydroxymethylation levels would be determined by subtracting the methylation levels of unoxidized and oxidized genomic DNA after bisulfite conversion, amplification and pyrosequencing.

I have observed no significant changes in methylation level between oxidized and unoxidized DNA from dCas9-TET1 transfected cells, but these findings must be taken with precaution because applied method has many possibilities for improvement and requires additional optimization. For example, I have used commercially available $KRuO_4$ from Alfa Aesar as described by Booth and collaborators (122), but the authors noted in the later work that they would recommend using a formulated $KRuO_4$ solution made by Cambridge Epigenetix for improved consistency and reproducibility in oxBS-seq (161). Also, 5fC does not convert to U as quickly as C does, so longer incubation times for bisulfite treatment might be required. I did not test different extended bisulfite treatments for dCas9-TET1 samples. Most striking issue I encountered in this analysis were differences in methylation levels for certain technical replicas (PCR products amplified from the same sample) between pyrosequencing runs on oxidized samples, which might imply that samples were not adequately purified between different steps of the oxBS procedure or that the oxidation

procedure resulted in sample degradation in some extent. I had issues with obtaining appropriate controls for this experiment which would confirm successful 5hmC to 5fC conversion and its detection. For further development of this method for dCas9-TET1 5hmC production assessment, I would consider implementing both oxidation and pyrosequencing controls in form of synthetic DNA substrates containing 5mc, 5hmC and C, which would be subjected to oxidation followed by bisulfite treatment and pyrosequencing. No detected methylation difference after oxBS between untransfected cells can be expected, considering that basal 5hmC levels in HEK293 cells are very low (113). Residue 5hmCs are known to be further oxidized to 5fC and 5caC residues by TET enzymes. Any difference in methylation levels between oxidized and unoxidized samples from dCas9-TET1 transfected cells would only be detectable if 5hmC generated by TET1 CD were present in such form in the moment of DNA isolation and not already further oxidized. It would be interesting to check for 5hmC production in transfected cells with a DNA dot blot assay at different time points after dCas9-TET1 transfection.

Off-target activity is a major concern when considering CRISPR/Cas9-based tools for clinical applications. Fu and collaborators found that off-target CRISPR/Cas9 cleavage activity can occur in DNA sequence with up to 5 mismatches in the 5' part of the sgRNA, and therefore imperfect pairing of CRISPR/Cas9 systems with endogenous loci is clearly a huge targeting issue (162). It is important to think of CRISPR/Cas9 DNA binding dynamics when analysing possible factors which might contribute to off-target activity. When expressed, Cas9 and the sgRNA form a RNP complex for which sgRNA scaffold folding is responsible. Cas9 changes its conformation upon sgRNA binding and takes an active DNA-binding conformation. Guide region of the sgRNA in such complex conformation is free to interact with its target sequence (10,59,163). The PAM sequence recognition is essential for the initiation of the Cas9 binding, but homology of the so called "seed*"* sequence in the sgRNA is critical for subsequent Cas9 binding and activity (10,59). Eight to 10 bases at the 3' end of sgRNA are considered to be the sgRNA "seed" sequence. After the "seed" sequence anneals to its target, sgRNA will continue to bind its target toward its 5' end. Such binding dynamics might explain why off-target activity is observed for sgRNAs with up to 5 mismatches outside their seed sequences. Strategies for reducing the off-target activity by sgRNA shortening have been employed. Fu and collaborators constructed a series of truncated sgRNAs (15, 16, 17, 18

and 19 bases) and validated their on-target and off-target activity. The sgRNAs of 15 nt to 16 nt had significantly decreased or undetectable activity, but sgRNAs of 17 and 18 nt had comparable on-target activity with classic 20 nt sgRNAs and exhibited less off-target activity (164). They reasoned that tolerating 5' mismatches in the sgRNA sequence might be a beneficial mechanism in naturally occurring CRISPR/Cas systems to tolerate the introduction of alternations in the target sequence. With dCas9-TET1, it is important to consider the catalytic domain activity as a contributing factor to off-target demethylation as well as the dCas9 binding properties. Overexpression of TET1 CD might be responsible for unwanted off-target activity; therefore, titration of the plasmid amount was done in this study to help minimize such off-target effect.

One strategy for reducing the off-target activity used in genome editing applications of CRISPR/Cas9 relies on using two sgRNAs targeting adjacent sites on opposite DNA strands, each guiding a Cas9 nickase variant (Cas9n; Cas9 with D10A mutation only). It has been demonstrated that paired nickase targeting can reduce off-target cleavage activity by 50- to 1500-fold (165). However, this strategy cannot be used for dealing with the off-target activity of dCas9-based tools as epigenome editing appears to be a much more complex approach. Wu and collaborators generated mouse ESCs with a stably integrated vector encoding HA-tagged dCas9 and performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) after transfections with several sgRNAs to create a genome-wide binding map of dCas9. They found that "seed" sequence as short as 5 bases followed by 5'-NGG-3' is sufficient for dCas9 binding, although less efficient than binding to the on-target site. They defined chromatin accessibility as the major determinant of dCas9 binding *in vivo*, which suggests that most of the dCas9 off-target binding might occur at active genes (163). They defined chromatin accessibility as the major determinant of *in vivo* dCas9 binding, which suggests that most of the dCas9 off-target binding might occur at active genes (163). Kim and collaborators demonstrated that delivery of the purified Cas9 protein and sgRNA not only circumvented the issue of transfection for hard-to-transfect fibroblasts and pluripotent stem cells, but also reduced the off-target cleavage activity (61). Using the RNP complexes as delivery method might be a nice way to temporally limit off-target activity because the RNPs are rapidly degraded in cells. However, this approach cannot be used for dCas9-based tools which have to be expressed for a longer period of time. Perhaps delivery of several epigenome editing

RNP complexes, inducing distinctive, synergistic epigenetic modifications at the single target site, might provide stability of the induced change in expression of target gene, so quick RNP degradation won't be a problem in such case and might even be beneficial for reducing the off-target effect.

Cas9 variants with broad PAM recognition and higher DNA specificity were developed to improve versatility and reduce off-target binding in a study performed by Hu and collaborators (166). This was achieved with Phage-assisted continuous evolution (PACE) system which enables hundreds of generations of directed evolution to occur weekly. Cas9 variants that accept an expanded range of PAM sequences, named xCas9s, were generated in this system under selective pressure to recognize different PAMs and maintain compatibility with different target DNA sequences. Generated xCas9 variants were analysed in contexts of transcriptional activation, genomic DNA cutting and base editing, and certain variants (xCas9 3.6 and xCas9 3.7) outperformed the SpCas9-based tools. Improved specificity of xCas9 3.6 and xCas9 3.7 was confirmed by genome-wide unbiased identification of DSBs enabled by sequencing (GUIDE-seq) after targeting of five endogenous genomic loci in HEK293T cells and two endogenous genomic loci in U2OS cells. Off-target modification was reduced by 43-fold in comparison with "classic" SpCas9. Also, transcriptional activation of endogenous genes using dxCas9-VPR tool was more precise and potent in comparison with classic dCas9-VPR-mediated activation, showing that dxCas9-VPR has compatible architecture with dCas9-based tools and can easily be used for gene activation at a substantially expanded set of PAMs (166). Using xCas9 in fusions with TET1 CD might be a nice way to reduce the observed off-target activity.

In this study, efforts have been made in dCas9-TET1 and TET1-dCas9 construction and validation to help minimize potential off-target effect. Chosen sgRNAs were validated with before mentioned GT Scan web-based tool and showed high specificity scores and initial dose-response experiment helped to determine optimal construct DNA amount with minimal off-target effect. Off-target activity of dCas9-TET1 was analysed by bisulfite pyrosequencing of *MGAT3* and *LAMB1* promoter regions in cells transfected with dCas9-TET1 and a non-targeting sgRNA. Some of the analysed CpG sites were significantly demethylated in such conditions and conclusion from these experiments would be that off-target activity of dCas9-TET1 is still a major concern. During the time-course experiment, off-target activity of

dCas9-TET1 was assessed by methylation analysis of CpG sites covered by MGAT3-A2 assay for cells transfected with dCas9-TET1-LAMB1-sg01 (Figure 24) and by methylation analysis of CpG sites covered by the LAMB1-A1 assay for cells transfected with dCas9-TET1-MGAT3-sg03 (Figure 25). Relatively high off-target activity was observed at certain CpG sites at certain time points. Since I have only analysed methylation levels of small fragments in promoter regions of these two genes (111 bp for *MGAT3* and 38 bp for *LAMB1* promoter) there is a possibility that dCas9-TET1 demethylates some other distal DNA regions, which come in proximity to the targeted region due to chromatin looping in a 3D DNA structure. For precise assessment of dCas9-TET1 on-target and off-target activity profiles, whole genome bisulfite sequencing (WGBS) or reduced representation bisulfite sequencing (RRBS) analyses should be done.

Targeted dCas9-based epigenome editing already greatly helps in elucidating causal relationships between epigenetic marks and genome regulation, with novel dCas9-based tools rapidly emerging with new potential applications. An ongoing effort is to apply the constructed dCas9-TET1 tools in biologically relevant contexts and to estimate induced physiologically relevant effects. Genome-wide association studies (GWAS) have identified genetic variants in the *MGAT3* and *LAMB1* loci to be associated with both IgG glycosylation and inflammatory bowel disease (23). *MGAT3* encodes the 1-4 N-acetylglucosaminyl transferase III, which adds bisecting N-acetyglucosamins (GlcNAcs) to N-glycans and *LAMB1* encodes the beta chain isoform laminin beta 1, which is one of the chains constituting the extracellular matrix glycoprotein laminin 1. Klasić and collaborators proposed that aberrant methylation of key glycosylation genes might lead to an increase in pro-inflammatory properties of IgG through aberrant glycosylation (18). They analysed methylation level in the *MGAT3* promoter from peripheral blood of several hundred IBD patients and healthy controls from two independent cohorts and found that differentially methylated CpG sites were hypermethylated in IBD patients compared to healthy individuals. The *LAMB1* gene was excluded from further methylation analysis in their study because overall methylation levels within *LAMB1* promoter were too low in both groups to detect differential methylation. They also suggested a direct functional role of *MGAT3* in IBD pathogenesis (18). One of the goals scheduled in this doctoral thesis was optimisation of delivery of constructed molecular tools to suspension cell line HEK293 FreeStyle, which

stably produce IgG, and to primary B-lymphocyte culture. However, optimization turned out to be an incredibly tedious task and the scope of these experiments surpassed the point of this research. Nevertheless, this stays a future goal for dCas9-TET1 and TET1-dCas9 application because those cells would be excellent biologically relevant models in studies of epigenetic modulation of endogenous *MGAT3* and *LAMB1* genes and direct analysis of subsequent IgG glycan phenotype. Also, one of the interesting future applications for dCas9-TET1 and TET1-dCas9 molecular tools might be targeted demethylation of enhancers and other DNA regulatory elements. It would also be interesting to see how would dCas9-TET1 and TET1-dCas9 perform in demethylation outside of the CpG rich context such as CGIs.

The dCas9-TET1 and TET1-dCas9 tools could have a therapeutic use for re-activation of targeted genes, for example tumor suppressors silenced by aberrant promoter methylation. The dCas9-TET1 and similar epigenome editing tools have a potential to replace epigenetic drugs of which several have been approved by the Food and Drug Administration Agency (FDA) and European Medicines Agency (EMA) for cancer treatment. For example, decitabine (5-aza-2-deoxycytidine) and azacitidine (5-azacytidine, Vidaza) are FDA approved non-specific methylation inhibitors in use for demethylation and re-expression of target genes in myelodysplastic syndromes and it would be great to replace them with much more specific tools for targeted demethylation (167,168). However, there are still issues to be solved in order to use CRISPR/dCas9 tools as epigenetic therapeutics, such as their off-target activity and their safe delivery to specific tissues (169).

6. Conclusions

- dCas9-TET1 molecular tool can be used for successful demethylation of targeted CpG sites.
- Different sgRNAs designed for *MGAT3* promoter targeting induced variable demethylation in CpG sites, depending on the sgRNA binding sites.
- Neighbouring CpG sites targeted with the same sgRNA differed in demethylation efficiency.
- Peak demethylation activity of the dCas9-TET1 was observed ≈26 bp downstream of sgRNA PAM recognition sequence, but it also demethylated proximal CpG sites upstream of PAM.
- Highest dCa9-TET1-induced methylation change observed at a single CpG site was 68% when targeted with a single sgRNA and 77% when targeted with a pool of 5 sgRNAs.
- TET1 catalytic domain was successfully inserted into a backbone for modular epitoolbox and the N-terminal TET1-dCas9 fusion was generated.
- The N-terminal TET1-dCas9 fusions exhibited a stronger demethylation activity both downstream and upstream from PAM recognition sequence, as well as stronger off-target activity than C-terminal dCas9-TET1 constructs.
- Highest methylation change induced by TET1-dCas9 was 87.1% at the CpG site 35 bp downstream from PAM.
- Peak demethylation activity was observed at 6^{th} and 7^{th} day after transfection for *MGAT3* and at 8^{th} day for *LAMB1* promoter, with steady gradual increase in methylation after 10^{th} day.
- The dCas9-TET1-induced CpG demethylation was not stable through cell divisions.
- Targeted demethylation of the *MGAT3* promoter with the dCas9-TET1 resulted in a 5-fold increase in *MGAT3* gene expression.
- VPR activation domain was successfully inserted into a backbone for assembly of modular epitoolbox and the VPR-dCas9 tool for CRISPRa was generated.
- Modest *MGAT3* gene activation effect (2-fold) was achieved with VPR-dCas9 targeting.
- Synergistic demethylation effect was observed for dCas9-TET1 simultaneously targeted with pooled 5 sgRNAs.
- Synergistic activation effect was observed for VPR-dCas9 simultaneously targeted with pooled 3 sgRNAs.
- Off-target activity of the dCas9-TET1 and TET1-dCas9 is a major concern and it should be analysed by WGBS or RRBS.

7. References

1. van Driel R, Fransz P, Verschure P. The eukaryotic genome: a system regulated at different hierarchical levels. J Cell Sci. 2003;115(20):4067–75.

2. Reid G, Gallais R, Metlvier R. Marking time: the dynamic role of chromatin and covalent modification in transcription. Int J Biochem Cell Biol. 2009;41(1):155–63.

3. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet. 2003;33(3):245–54.

4. Almouzni G, Cedar H. Maintenance of Epigenetic Information. Cold Spring Harb Perspect Biol. 2016;8(5):a019372.

5. Skinner MK. Role of epigenetics in developmental biology and transgenerational inheritance. Birth Defects Res C Embryo Today. 2011;93(1):51–5.

6. Kim HJ, Lee HJ, Kim H, Cho SW, Kim J-S. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. Genome Res. 2009 Jul;19(7):1279–88.

7. Wilson KA, McEwen AE, Pruett-Miller SM, Zhang J, Kildebeck EJ, Porteus MH. Expanding the repertoire of target sites for zinc finger nuclease-mediated genome modification. Mol Ther - Nucleic Acids. 2013;2(4):e88.

8. Boch J. TALEs of genome targeting. Nat Biotechnol. 2011;29(2):135–6.

9. Morbitzer R, Elsaesser J, Hausner J, Lahaye T. Assembly of custom TALE-type DNA binding domains by modular cloning. Nucleic Acids Res. 2011;39(13):5790–9.

10. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816–21.

11. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc. 2013;8(11):2281–308.

12. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, et al. CRISPR-

mediated modular RNA-guided regulation of transcription in eukaryotes. Cell. 2013;154(2):442–51.

13. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol. 2015;33(5):510–7.

14. Vojta A, Dobrinic P, Tadic V, Bockor L, Korac P, Julg B, et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res. 2016;44(12):5615–28.

15. Robertson KD. DNA methylation and human disease. Nat Rev Genet. 2005;6(8):597–610.

16. Horvat T, Zoldoš V, Lauc G. Evolutional and clinical implications of the epigenetic regulation of protein glycosylation. Clin Epigenetics. 2011;2(2):425–32.

17. Vojta A, Dobrinic P, Tadic V, Bockor L, Korac P, Julg B, et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic Acids Res. 2016;44(12):5615–28.

18. Klasić M, Markulin D, Vojta A, Samaržija I, Biruš I, Dobrinić P, et al. Promoter methylation of the MGAT3 and BACH2 genes correlates with the composition of the immunoglobulin G glycome in inflammatory bowel disease. Clin Epigenetics. 2018;10(1):75.

19. Liu D, Chu X, Wang H, Dong J, Ge S-Q, Zhao Z-Y, et al. The changes of immunoglobulin G N-glycosylation in blood lipids and dyslipidaemia. J Transl Med. 2018;16(1):235.

20. Klose RJ, Bird AP. Genomic DNA methylation: The mark and its mediators. Trends Biochem Sci. 2006;31(2):89–97.

21. Tahiliani M, Koh KP, Shen Y, Pastor W a, Brudno Y, Agarwal S, et al. Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. Science. 2009;324(5929):930–5.

22. Ooi SKT, Bestor TH. The Colorful History of Active DNA Demethylation. Cell.

2008;133(7):1145–8.

23. Lauc G, Huffman JE, Pučić M, Zgaga L, Adamczyk B, Mužinić A, et al. Loci Associated with N-Glycosylation of Human Immunoglobulin G Show Pleiotropy with Autoimmune Diseases and Haematological Cancers. PLoS Genet. 2013;9(1):e1003225.

24. Vink T, Oudshoorn-Dickmann M, Roza M, Reitsma J-J, de Jong RN. A simple, robust and highly efficient transient expression system for producing antibodies. Methods. 2014;65(1):5–10.

25. Kao S-H, Wu K-J, Lee W-H. Hypoxia, Epithelial-Mesenchymal Transition, and TET-Mediated Epigenetic Changes. J Clin Med. 2016;5(2):e24.

26. Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, P R Iyer E, et al. Highly efficient Cas9-mediated transcriptional programming. Nat Methods. 2015;12(4):326–8.

27. Danna K, Nathans D. Specific cleavage of simian virus 40 DNA by restriction endonuclease of Hemophilus influenzae. Proc Natl Acad Sci. 1971;68(12):2913–7.

28. Smith HO, Welcox KW. A Restriction enzyme from Hemophilus influenzae I. J Mol Biol. 1970;51(2):379–91.

29. Kim H, Kim JS. A guide to genome engineering with programmable nucleases. Nat Rev Genet. 2014;15(5):321–34.

30. Thomas KR, Folger KR, Capecchi MR. High frequency targeting of genes to specific sites in the mammalian genome. Cell. 1986;44(3):419–28.

31. Doetschman T, Gregg RG, Maeda N, Hooper ML, Melton DW, Thompson S, et al. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. Nature. 1987;330(6148):576–8.

32. Bibikova M, Golic M, Golic KG, Carroll D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics. 2002;161(3):1169–75.

33. Adli M. The CRISPR tool kit for genome editing and beyond. Nat Commun. 2018;9(1):1911.

34. Krueger U, Bergauer T, Kaufmann B, Wolter I, Pilk S, Heider-Fabian M, et al. Insights into Effective RNAi Gained from Large-Scale siRNA Validation Screening. Oligonucleotides. 2007;17(2):237–50.

35. Jasin M. Genetic manipulation of genomes with rare-cutting endonucleases. Trends Genet. 1996;12(6):224–8.

36. Chames P, Epinat J-C, Guillier S, Patin A, Lacroix E, Pâques F. In vivo selection of engineered homing endonucleases using double-strand break induced homologous recombination. Nucleic Acids Res. 2005;33(20):e178–e178.

37. Hanas JS, Hazuda DJ, Bogenhagen DF, Wu FYH, Wu CW. Xenopus Transcription Factor-a Requires Zinc for Specific Binding to the 5s-Rna Gene. Fed Proc. 1983;42(7):1881.

38. Pavletich NP, Pabo C. Zinc Structure of a Recognition : Complex Zif268-DNA. Adv Sci. 1991;252(5007):809–17.

39. Kim YG, Cha J, Chandrasegaran S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc Natl Acad Sci. 1996;93(3):1156–60.

40. Smith J, Bibikova M, Whitby FG, Reddy AR, Chandrasegaran S, Carroll D. Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. Nucleic Acids Res. 2000;28(17):3361–9.

41. Dreier B, Fuller RP, Segal DJ, Lund C V., Blancafort P, Huber A, et al. Development of zinc finger domains for recognition of the 5′-CNN-3′ family DNA sequences and their use in the construction of artificial transcription factors. J Biol Chem. 2005;280(42):35588–97.

42. Ramirez CL, Foley JE, Wright DA, Müller-Lerch F, Rahman SH, Cornu TI, et al. Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods. 2008;5(5):374–5.

43. Gutschner T, Baas M, Diederichs S. Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. Genome Res. 2011;21(11):1944–54.

44. Cathomen T, Keith Joung J. Zinc-finger nucleases: The next generation emerges. Mol Ther. 2008;16(7):1200–7.

45. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, et al. Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. Genetics. 2010;186(2):757–61.

46. Boch J, Bonas U. Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. Annu Rev Phytopathol. 2010;48(1):419–36.

47. Bonas U, Stall RE, Staskawicz B. Genetic and structural characterization of the avirulence gene avrBs3 from Xanthomonas campestris pv. vesicatoria. MGG Mol Gen Genet. 1989;218(1):127–36.

48. Zhang Y, Zhang F, Li X, Baller JA, Qi Y, Starker CG, et al. Transcription Activator-Like Effector Nucleases Enable Efficient Plant Genome Engineering. Plant Physiol. 2013;161(1):20–7.

49. Moscou MJ, Bogdanove AJ. A Simple Chiper Governs DNA Recognition by TAL Effectors. Science. 2009;326(5959):1501.

50. Kim Y, Kweon J, Kim A, Chon JK, Yoo JY, Kim HJ, et al. A library of TAL effector nucleases spanning the human genome. Nat Biotechnol. 2013;31(3):251–8.

51. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 2012;482(7385):331–8.

52. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol. 2005;60(2):174–82.

53. Bolotin A, Quinquis B, Sorokin A, Dusko Ehrlich S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005;151(8):2551–61.

54. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science. 2007;315(3):1709–13.

55. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol. 2015;13(11):722–36.

56. Makarova KS, Zhang F, Koonin E V. SnapShot: Class 2 CRISPR-Cas Systems. Cell. 2017;168(1–2):328–328.e1.

57. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 2011;471(7340):602–7.

58. Marraffini LA. The CRISPR-Cas system of Streptococcus pyogenes : function and applications. In: Streptococcus pyogenes : Basic Biology to Clinical Manifestations [Internet] University of Oklahoma Health Sciences Center. 2016. p. 1–13.

59. Cong L, Ran FA, Cox D, Lin S, Barretto R, Hsu PD, et al. Multiplex Genome Engineering Using CRISPR/VCas Systems. Science. 2013;339(6121):819–23.

60. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. Science. 2013;339(6121):823–6.

61. Kim S, Kim D, Cho SW, Kim J, Kim J-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 2014;24(6):1012–9.

62. Pulecio J, Verma N, Mejía-Ramírez E, Huangfu D, Raya A. CRISPR/Cas9-Based Engineering of the Epigenome. Cell Stem Cell. 2017;21(4):431–47.

63. Keung AJ, Bashor CJ, Kiriakov S, Collins JJ, Khalil AS. Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation. Cell. 2014;158(1):110–20.

64. Rivenbark AG, Stolzenburg S, Beltran AS, Yuan X, Rots MG, Strahl BD, et al. Epigenetic reprogramming of cancer cells via targeted DNA methylation. Epigenetics. 2012;7(4):350–60.

65. Nunna S, Reinhardt R, Ragozin S, Jeltsch A. Targeted methylation of the epithelial cell adhesion molecule (EpCAM) promoter to silence its expression in ovarian cancer cells.

PLoS One. 2014;9(1):e87703–e87703.

66. Zhang B, Xiang S, Zhong Q, Yin Y, Gu L, Deng D. The p16-specific reactivation and inhibition of cell migration through demethylation of CpG islands by engineered transcription factors. Hum Gene Ther. 2012;23(10):1071–81.

67. Maeder ML, Linder SJ, Reyon D, Angstman JF, Fu Y, Sander JD, et al. Robust, synergistic regulation of human gene expression using TALE activators. Nat Methods. 2013;10(3):243–5.

68. Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol. 2011;29(2):149–53.

69. Maeder ML, Angstman JF, Richardson ME, Linder SJ, Cascio VM, Tsai SQ, et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. Nat Biotechnol. 2013;31(12):1137–42.

70. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Methods. 2013;10(10):973–6.

71. Stepper P, Kungulovski G, Jurkowska RZ, Chandra T, Krueger F, Reinhardt R, et al. Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. Nucleic Acids Res. 2017;45(4):1703–13.

72. Mendenhall EM, Williamson KE, Reyon D, Zou JY, Ram O, Joung JK, et al. Locus-specific editing of histone modifications at endogenous enhancers. Nat Biotechnol. 2013;31(12):1133–6.

73. Shechner DM, Hacisuleyman E, Younger ST, Rinn JL. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nat Methods. 2015;12(7):664–70.

74. Waddington CH. The epigenotype. Endeavour. 1942;1:18–20.

75. Holliday R, Pugh J. DNA modification mechanisms and gene activity during development. Science. 1975;187(4173):226–32.

76. Riggs AD. X inactivation, differentiation and DNA methylation. Cytogenet Cell Genet. 1975;14:9–25.

77. Spencer V, Davie J. Role of covalent modifications of histones in regulating gene expression. Gene. 1999;240(1):1–12.

78. Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, et al. CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development. 2009 Sep;136(18):3131–41.

79. Holoch D, Moazed D. RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet. 2015;16(2):71–84.

80. Meister G, Landthaler M, Patkaniowska A, Dorsett Y, Teng G, Tuschi T. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell. 2004;15(2):185–97.

81. Consortium EP. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004;306(5696):636–40.

82. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799–816.

83. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010;11(3):204–20.

84. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315–22.

85. Bird AP. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 1980;8(7):1499–504.

86. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr ARW, James KD, Turner DJ, et al. Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome. PLoS Genet. 2010;6(9):e1001134.

87. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev. 2011;25(10):1010–22.

88. Nan X, Ng H, Johnson C, Laherty C, Turner B, Eisenman R, et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature. 1998;393(6683):386–9.

89. Li Y, Xu Q, Lv N, Wang L, Zhao H, Wang X, et al. Clinical implications of genome-wide DNA methylation studies in acute myeloid leukemia. J Hematol Oncol. 2017;10(1):41.

90. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases Non-invasive sexing of preimplantation stage mammalian embryos. Nat Am Inc. 1998;19(7):219–20.

91. Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. J Biol Chem. 2004;279(46):48350–9.

92. Vandiver AR, Idrizi A, Rizzardi L, Feinberg AP, Hansen KD. DNA methylation is stable during replication and cell cycle arrest. Sci Rep. 2015;5:17911.

93. Okano M, Bell DW, Haber DA, Li E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. Cell. 1999;99:247–57.

94. Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007;448(7154):714–7.

95. Jeltsch A, Ehrenhofer-Murray A, Jurkowski TP, Lyko F, Reuter G, Ankri S, et al. Mechanism and biological role of Dnmt2 in Nucleic Acid Methylation. RNA Biol. 2017;14(9):1108–23.

96. Nguyen P, Cui H, Bisht KS, Sun L, Patel K, Lee RS, et al. CTCFL/BORIS Is a Methylation-Independent DNA-Binding Protein That Preferentially Binds to the Paternal H19 Differentially Methylated Region. Cancer Res. 2008;68(14):5546–51.

97. Guastafierro T, Cecchinelli B, Zampieri M, Reale A, Riggio G, Sthandier O, et al. CCCTC-binding Factor Activates PARP-1 Affecting DNA Methylation Machinery. J Biol Chem. 2008;283(32):21873–80.

98. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5900):517–20.

99. Morgan HD, Santos F, Green K, Dean W, Reik W. Epigenetic reprogramming in mammals. Hum Mol Genet. 2005;14(1):47–58.

100. Li E. Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet. 2002;3(9):662–73.

101. Avner P, Heard E. X-chromosome inactivation: counting, choice and initiation. Nat Rev Genet. 2001;2(1):59–67.

102. Walsh CP, Chaillet JR, Bestor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet. 1998;20(2):116–7.

103. McCabe MT, Brandes JC, Vertino PM. Cancer DNA methylation: molecular mechanisms and clinical implications. Clin Cancer Res. 2009;15(12):3927–37.

104. Calvello M, Tabano S, Colapietro P, Maitz S, Pansa A, Augello C, et al. Quantitative DNA methylation analysis improves epigenotype-phenotype correlations in Beckwith-Wiedemann syndrome. Epigenetics. 2013;8(10):1053–60.

105. Muralidhar B, Butler MG. Methylation PCR Analysis of Prader-Willi Syndrome, Angelman Syndrome, and Control Subjects. Am J Med Genet. 1998;80(3):263–5.

106. Amir R, Van den Veyver I, Wan M, Tran C, Francke U, Zoghbi H. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet. 1999;23(2):185–8.

107. Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell. 2014;156(1–2):45–68.

108. Bruniquel D, Schwartz RH. Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nat Immunol. 2003;4(3):235–40.

109. Gjerset RA, Martin DW. Presence of a DNA Demethylating Activity in the Nucleus of Murine Erythroleukemic Cells. J Biol Chem. 1982;257(15):8581–3.

110. Bhattacharya S, Ramchandani S, Cervoni N, Szyf M. A mammalian protein with specific demethylase activity for mCpG DNA. Nature. 1999;397(6720):579–83.

111. Detich N, Theberge J, Szyf M. Promoter-specific activation and demethylation by MBD2/demethylase. J Biol Chem. 2002;277(39):35791–4.

112. Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska R, Carmouche R, et al. Cyclical DNA methylation of a transcriptionally active promoter. Nature. 2008;452(7183):45–50.

113. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300–3.

114. Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. Nature. 2013;502(7472):472–9.

115. Loenarz C, Schofield CJ. Oxygenase Catalyzed 5-Methylcytosine Hydroxylation. Chem Biol. 2009;16(6):580–3.

116. Borst P, Sabatini R. Base J: discovery, biosynthesis and possible functions. Annu Rev Microbiol. 2008;62:235–51.

117. Lorsbach R, Moore J, Mathew S, Raimondi S, Mukatira S, Downing J. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia. 2003;17(3):637–41.

118. Franchini D-M, Schmitz K-M, Petersen-Mahrt SK. 5-Methylcytosine DNA Demethylation: More Than Losing a Methyl Group. Annu Rev Genet. 2012;46(1):419–41.

119. Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol. 2013 Jun;14(6):341–56.

120. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development,

and cancer. Genes Dev. 2016;30(7):733–50.

121. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. PLoS One. 2010 Jan 26;5(1):e8888.

122. Booth MJ, Branco MR, Ficz G, Oxley D, Krueger F. Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. Science. 2012;8333(May):934–8.

123. Wyatt GR, Cohen SS. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. Biochem J. 1953;55(5):774–82.

124. Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. Nat Chem. 2014;6:1049–55.

125. Kriaucionis S, Heintz N. The nuclear DNA base, 5-hydroxymethylcytosine is present in brain and enriched in Purkinje neurons. Science. 2009;324(5929):929–30.

126. Rudenko A, Dawlaty MM, Seo J, Cheng AW, Meng J, Le T, et al. Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. Neuron. 2013;79(6):1109—1122.

127. Green BB, Houseman EA, Johnson KC, Guerin DJ, Armstrong DA, Christensen BC, et al. Hydroxymethylation is uniquely distributed within term placenta, and is associated with gene expression. FASEB J. 2016;30(8):2874–84.

128. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. 2011;12(6):R54.

129. Koh KP, Yabuuchi A, Rao S, Huang Y, Cunniff K, Laiho A, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell. 2012;8(2):200–13.

130. Ito S, Alessio ACD, Taranova O V, Hong K, Lawrence C. Role of Tet proteins in 5mC to 5hmC conversion , ES cell self- renewal , and ICM specification. Nature. 2010;466(7310):1129–33.

131. Dawlaty MM, Ganz K, Powell BE, Hu Y-C, Markoulaki S, Cheng AW, et al. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell. 2011;9(2):166–75.

132. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. 2010;38(19):e181–e181.

133. Dawlaty MM, Breiling A, Le T, Barrasa MI, Raddatz G, Gao Q, et al. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Dev Cell. 2014;29(1):102–11.

134. Tsagaratou A, Äijö T, Lio C-WJ, Yue X, Huang Y, Jacobsen SE, et al. Dissecting the dynamic changes of 5-hydroxymethylcytosine in T-cell development and differentiation. Proc Natl Acad Sci. 2014;111(32):e3306–15.

135. Iurlaro M, Ficz G, Oxley D, Raiber E-A, Bachman M, Booth MJ, et al. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome Biol. 2013;14(10):R119.

136. Jin C, Lu Y, Jelinek J, Liang S, Estecio MRH, Barton MC, et al. TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. Nucleic Acids Res. 2014;42(11):6956–71.

137. Ficz G, Gribben JG. Loss of 5-hydroxymethylcytosine in cancer: cause or consequence? Genomics. 2014;104(5):352–7.

138. Li Z, Cai X, Cai C-L, Wang J, Zhang W, Petersen BE, et al. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood. 2011;118(17):4509–18.

139. Yanagawa N, Tamura G, Honda T, Endoh M, Nishizuka S, Motoyama T. Demethylation of the synuclein gamma gene CpG island in primary gastric cancers and gastric cancer cell lines. Clin Cancer Res. 2004;10(7):2447–51.

140. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods. 2015;11(8):783–4.

141. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 2014;32(4):347–55.

142. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods. 2001;25(4):402–8.

143. Chen X, Zaro JL, Shen W-C. Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013;65(10):1357–69.

144. Luke GA, de Felipe P, Lukashev A, Kallioinen SE, Bruno EA, Ryan MD. Occurrence, function and evolutionary origins of "2A-like" sequences in virus genomes. J Gen Virol. 2008;89(4):1036–42.

145. O'Brien A, Bailey TL. GT-Scan: identifying unique genomic targets. Bioinformatics. 2014;30(18):2673–5.

146. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. Nat Methods. 2013;10(10):977–9.

147. Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res. 2013;23(10):1163–71.

148. Radzisheuskaya A, Shlyueva D, Müller I, Helin K. Optimizing sgRNA position markedly improves the efficiency of CRISPR/dCas9-mediated transcriptional repression. Nucleic Acids Res. 2016;44(18):e141–e141.

149. Xu X, Tao Y, Gao X, Zhang L, Li X, Zou W, et al. A CRISPR-based approach for targeted DNA demethylation. Cell Discov. 2016;2:16009.

150. Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 2014;156(5):935–49.

151. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2015;517(7536):583–8.

152. Zoldoš V, Horvat T, Novokmet M, Cuenin C, Mužinić A, Pučić M, et al. Epigenetic silencing of HNF1A associates with changes in the composition of the human plasma N-glycome. Epigenetics. 2012;7(2):164–72.

153. Mabaera R, Richardson CA, Johnson K, Hsu M, Fiering S, Lowrey CH. Developmental- and differentiation-specific patterns of human gamma- and beta-globin promoter DNA methylation. Blood. 2007 Aug 15;110(4):1343–52.

154. Lo C-L, Choudhury SR, Irudayaraj J, Zhou FC. Epigenetic Editing of Ascl1 Gene in Neural Stem Cells by Optogenetics. Sci Rep. 2017;7:42047.

155. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev. 2002;1(16):6–21.

156. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc. 2013;8:2180–96.

157. Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol. 2013;31(3):230–2.

158. Lee CM, Cradick TJ, Bao G. The Neisseria meningitidis CRISPR-Cas9 System Enables Specific Genome Editing in Mammalian Cells. Mol Ther. 2016;24(3):645–54.

159. Kungulovski G, Nunna S, Thomas M, Zanger UM, Reinhardt R, Jeltsch A. Targeted epigenome editing of an endogenous locus with chromatin modifiers is not stably maintained. Epigenetics Chromatin. 2015;8:12.

160. Yu M, Hon GC, Szulwach KE, Song C-X, Jin P, Ren B, et al. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. Nat Protoc. 2012;7(12):2159–70.

161. Booth MJ, Ost TWB, Beraldi D, Bell NM, Branco MR, Reik W, et al. Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. Nat Protoc. 2013;8(10):1841–51.

162. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol. 2013;31(9):822–6.

163. Wu X, Scott DA, Kriz AJ, Chiu AC, Hsu PD, Dadon DB, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat Biotechnol. 2014;32(7):670–6.

164. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotechnol. 2014;32(3):279–84.

165. Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, Trevino AE, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell. 2013;154(6):1380–9.

166. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature. 2018;556(7699):57–63.

167. Marks PW. Decitabine for acute myeloid leukemia. Expert Rev Anticancer Ther. 2012;12(3):299–305.

168. Kimura S, Kuramoto K, Homan J, Naruoka H, Ego T, Nogawa M, et al. Antiproliferative and antitumor effects of azacitidine against the human myelodysplastic syndrome cell line SKM-1. Anticancer Res. 2012;32(3):795–8.

169. Jiang F, Doudna JA. CRISPR–Cas9 Structures and Mechanisms. Annu Rev Biophys. 2017;46(1):505–29.

8. List of abbreviations

AAV – adeno-associated virus

AGO – Argonaute family of proteins

ANOVA – analysis of variance

ATF – artificial transcription factor

BER – base excision repair

Cas9 – CRISPR associated protein 9

Cas9n – Cas9 nickase

CCR5 – C-C chemokine receptor type 5

CD – catalytic domain

CGI – CpG island

CMS – cytosine 5-methylenesulfonate

CpG – cytosine (C) followed by guanine (G) in DNA sequence, CpG dinucleotide

CRISPR/Cas –clustered regularly interspaced palindromic repeats/CRISPR associated genes

CRISPRa – CRISPR activation

CRISPRi – CRISPR interference

crRNA – mature CRISPR RNA transcript

dCas9 – dead Cas9; nuclease-deficient Cas9

DED1 - “dead” TET1; catalytically inactive variant of TET1

DNMT3A –DNA methyltransferase 3A

DNMT3B –DNA methyltransferase 3B

DNMT1 – DNA methyltransferase 1

DNMT3L –DNA methyltransferase 3-like

DNMT2 – DNA methyltransferase 2

DSB – double stranded break

DSBH – double stranded β-helix domain

EGFP – enhanced green fluorescent protein

EMA –European Medicines Agency

EMX1 – Empty Spiracles Homeobox 1

ENCODE- ENCyclopedia Of DNA Elements

EpCAM - epithelial cell adhesion molecule

ESCs – embryonic stem cells

FDA – Food and Drug Administration

GUIDE-seq – Genome-wide Unbiased IDentification of DSBs Enabled by sequencing

GWAS – genome wide association study

HDR – homology directed repair

HEK293 – human embryonic kidney 293 cells

HEK293FS – human embryonic kidney 293 FreeStyle cells

hmeDIP-seq – hydroxymethyl DNA immunoprecipitation followed by deep sequencing

HOMedU – hydroxymethyldeoxyuridine

HSF1 – human heat-shock factor 1

H3K4me0 – unmethylated histone H3 lysine 4

H3K27 – histone H3 lysine 27

H3K27ac – histone H3 lysine 27 acetylation

H3K27me3 – trimethylated histone H3 lysine 27

H3K9me –histone H3 lysine 9 methylation

IBD – inflammatory bowel disease

IgG – immunoglobulin G

IL1RN – interleukin-1 receptor antagonist

INDELs – small insertions/deletions

iPSCs – induced pluripotent stem cells

J – β-D-glucosyl hydroxyuracil

JBP1, JBP2 – thymine dioxygenase 1 and 2

Klf4 – Kruppel Like Factor 4

KRAB – Krüppel-associated box

K562 – human erythroleukemia cell line K562

VP16 - *Herpes simplex* virus protein vmw65

VP64 – VP16 tetramer

VPR – VP64-p65-Rta synthetic activator domain

lncRNA – long non-coding RNA

LAMB1 – Laminin Subunit Beta 1

MALAT1 – Metastasis Associated Lung Adenocarcinoma Transcript 1

Maspin – mammary serine protease inhibitor

MCF7 – human breast cancer cell line MCF7

MCP – MS2 bacteriophage coat protein

MS2-p65-HSF1 – activation domain from human heat-shock factor 1 fused to MCP

MGAT3 – Mannosyl (Beta-1,4-)-Glycoprotein Beta-1,4-N-Acetylglucosaminyltransferase

MBD – methyl-binding domain

MBD2 – methyl-binding domain 2

MeCP2 – methyl-CpG-binding protein 2

miRNA – microRNA

NHEJ – non-homologous end joining

NLS – nuclear localization signal

NmCas9 – *Neisseria meningitides* Cas9

Oct4 - octamer-binding transcription factor 4

oxBS-seq – oxidative bisulfite sequencing

PACE – Phage-Assisted Continuous Evolution

PAM – protospacer adjacent motif

PCR – polymerase chain reaction

PGCs – primordial germ cells

pre-crRNA – precursor CRISPR RNA

p300 - p300 transcriptional co-activator

qRT-PCR – quantitative reverse transcription polymerase chain reaction

RRBS – reduced representation bisulfite sequencing

RVD – repeat variable diresidue

SaCas9 – *Staphylococcus aureus* Cas9

SAH – S-adenosylhomocysteine

SAM - S-adenosylmethionine

sgRNA – single guide RNA

sgNT – non targeting sgRNA

siRNA – short interfering RNA

SKOV3 – human ovarian cancer cell line SKOV3

SNCG – synuclein gamma

SOX2 – (sex determining region Y)-box 2

SpCas9 – *Streptococcus pyogenes* Cas9

SUM159 – human breast cancer cell line SUM159

TAB-seq – Tet-assisted bisulfite sequencing

TALE – Transcription activator-like effector

TALEN – Transcription activator-like effector nuclease

TDG – thymine DNA glycosylase

TET1 – Ten-eleven translocation 1 protein

TF – transcription factor

tracrRNA – transactivating CRISPR RNA

TSS – transcription start site

VEGF-A – Vascular endothelial growth factor A

WGBS – whole genome bisulfite sequencing

Xic – X-chromosome-inactivation-centre

ZF – Zinc Finger

ZFN – Zinc Finger nuclease

2OG – 2-oxoglutarate

5mC – 5-methylcytosine

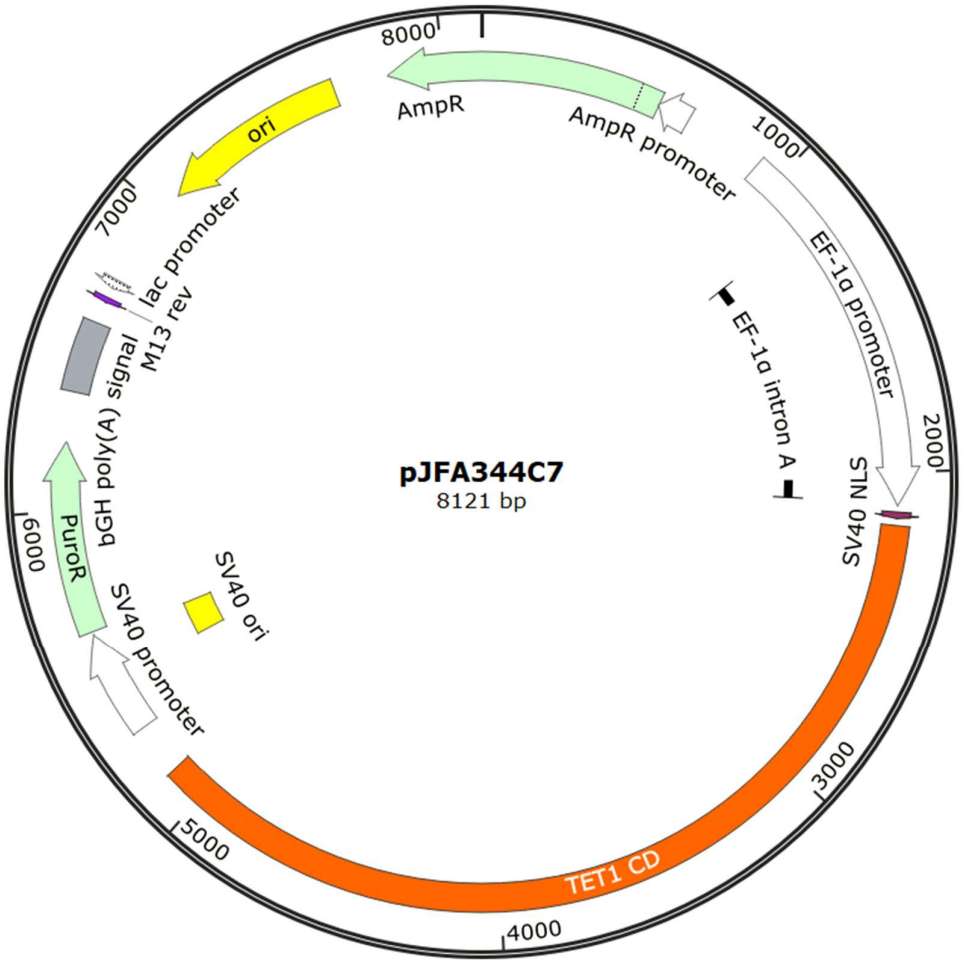
5hmC – 5-hydroxymethylcytosine

5fC – 5-formylcytosine

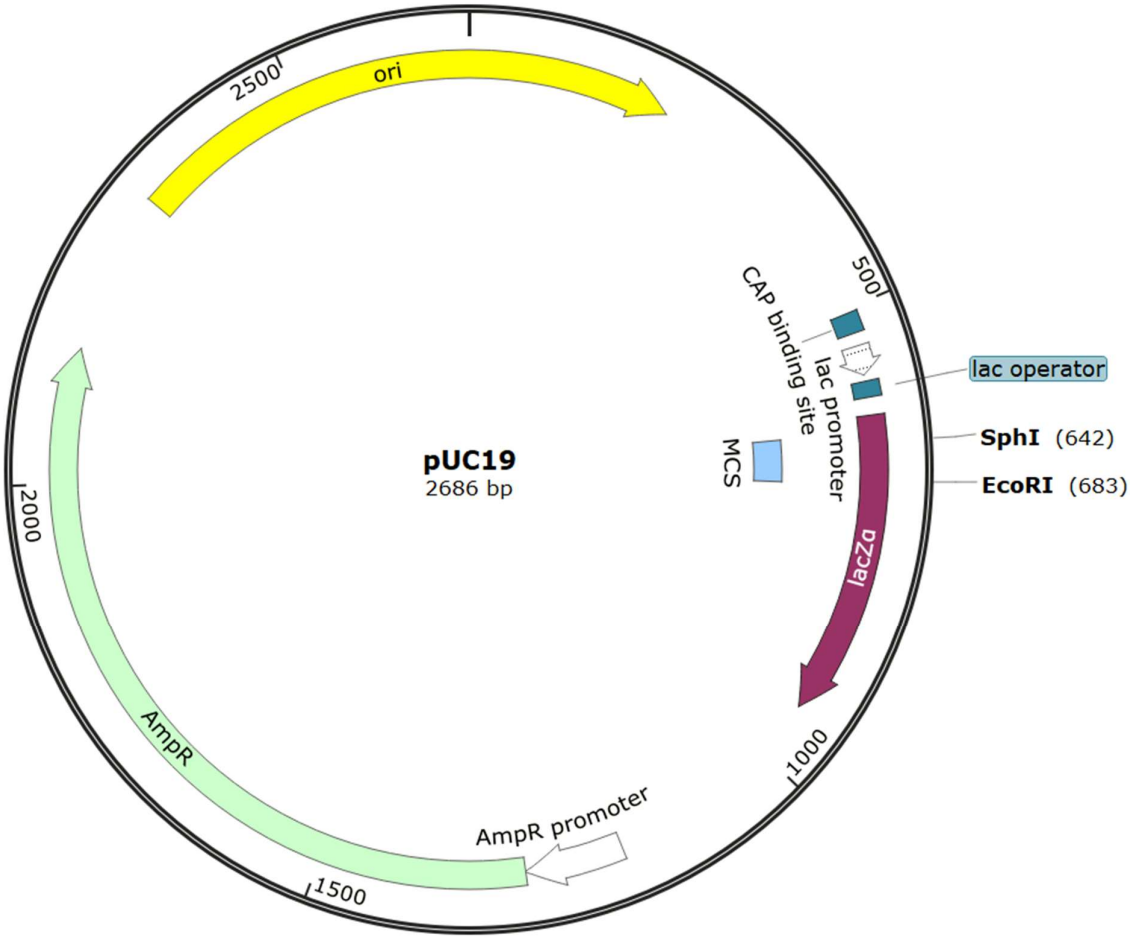
5caC – 5-carboxycytosine

9. Appendix 1

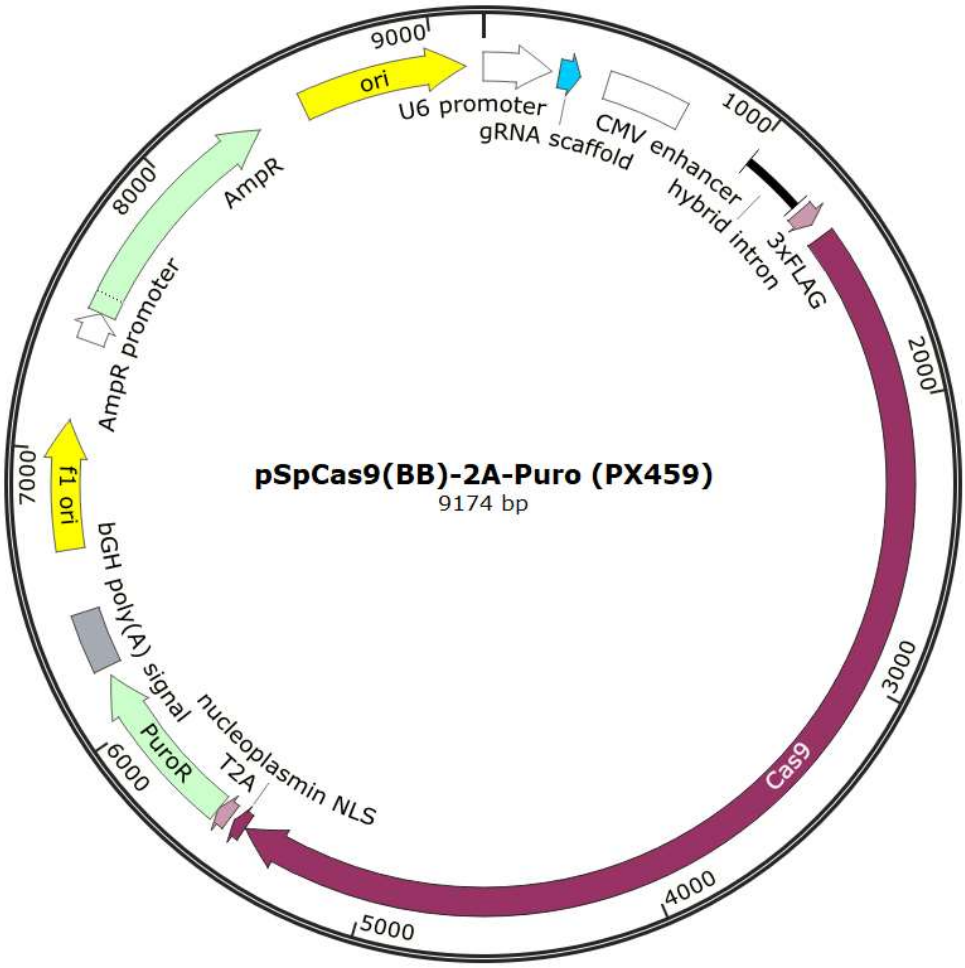
Maps of plasmids used for dCas9-TET1 construction



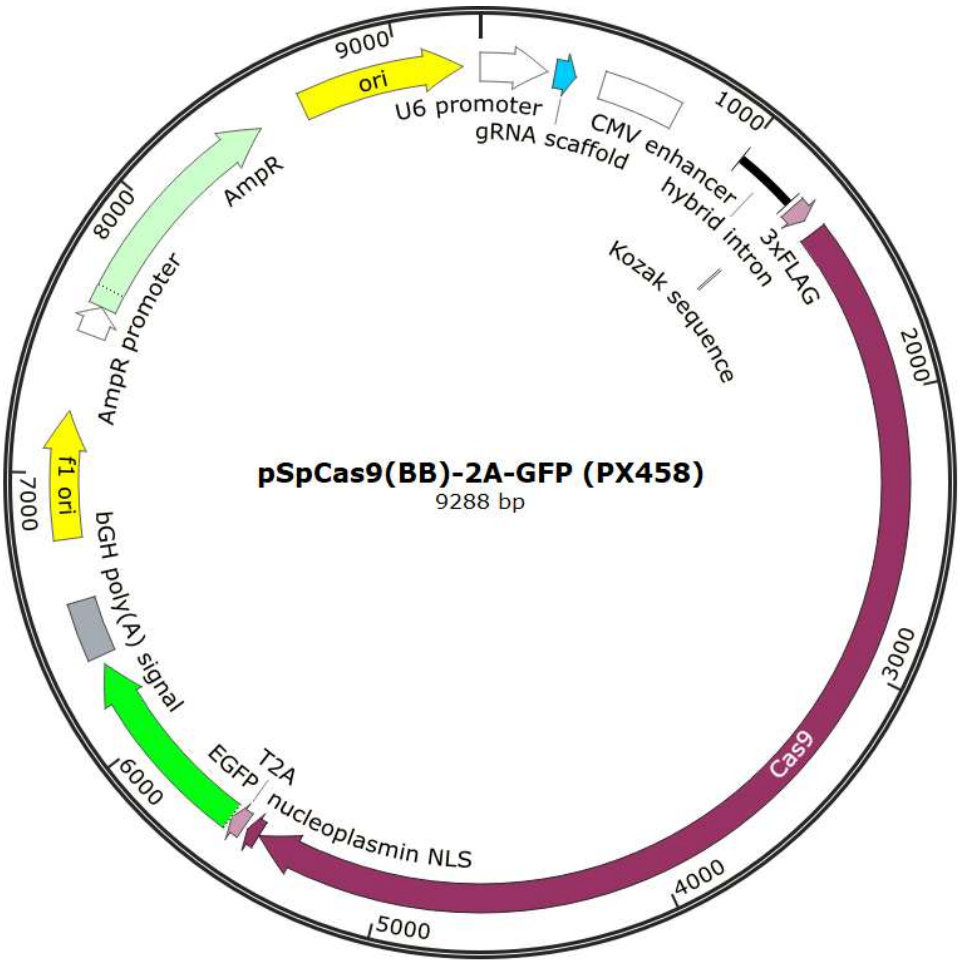
pJFA344C7. Addgene plasmid #49236. Used as source of TET1 CD.



pUC19. Addgene plasmid #50005. Used as shuttle vector for insertion of TET1 CD via indicated SphI and EcoRI sites and targeted mutageneses before subsequent cloning events.



pSpCas9n(BB)-2A-Puro. Addgene plasmid #48141. Used as source of puromycin resistance marker which was additionally mutated.



pSpCas9n(BB)-2A-GFP. Addgene plasmid #48140. Used as source of EGFP selection marker.

10. Appendix 2

Amino acid sequences of constructed fusion proteins

dCas9-TET1-PuroR (2376 aa):

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL
RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH
DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR
RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV
DQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN
YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR
MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL
ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE
VLDATLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKKLEGGGGSGSLPTCS
CLDRVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSH
GCPIAKWVLRRSSDEEKVLCLVRQRTGHHCPTAVMVVLIMVWDGIPLPMADRLYTE
LTENLKSYNGHPTDRRCTLNENRTCTCQGIDPETCGASFSFGCSWSMYFNGCKFGRS
PSPRRFRIDPSSPLHEKNLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLG
SKEGRPFSGVTACLDFCAHPHRDIHNMNNGSTVVCTLTREDNRSLGVIPQDEQLHVL
PLYKLSDTDEFGSKEGMEAKIKSGAIEVLAPRRKKRTCFTQPVPRSGKKRAAMMTEV
LAHKIRAVEKKPIPRIKRKNNSTTTNNSKPSSLPTLGSNTETVQPEVKSETEPHFILKSS
DNTKTYSLMPSAPHPVKEASPGFSWSPKTASATPAPLKNDATASCGFSERSSTPHCTM
PSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSEPSTGVTEPLTPHQPNHQ
PSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWSDSEHIFL
DANIGGVAIAPAHGSVLIECARRELHATTPVEHPNRNHPTRLSLVFYQHKNLNKPQH

GFELNKIKFEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQIPSHKALTLTHDN
VVTVSPYALTHVAGPYNHWVGRPEFGSGEGRGSLLTCGDVEENPGPMTEYKPTVR
LATRDDVPRAVRTLAAAFADYPATRHTVDPDRHIERVTELQELFLTRVGLDIGK
VWVADDGAAVAVWTTPESVEAGAVFAEIGPRMAELSGSRLAAQQQMEGLLAP
HRPKEPAWFLATVGVSPDHQGKGLGSAVVLPGVEAAERAGVPAFLETSAPRNLP
FYERLGFTVTADVEVPEGPRTWCMTRKPGA*

3xFLAG ; SV40 NLS ; dSpCas9 ; nucleoplasmin NLS ; **Gly4Ser linker** ; **TET1 CD** (substitutions present in DED1 variant are indicated; **H → Y** and **D → A)** ; T2A self-cleaving peptide ; PuroR ; STOP (*)

dCas9-TET1-eGFP (2420 aa):

MDYKDHDGDYKDHDIDYKDDDDKMAPKKKRKVGIHGVPAADKKYSIGLAIGTNSV
GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT
RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLV
QTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL
TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL
RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYID
GGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL
RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEV
VDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP
AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH
DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR
RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA
QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ
TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV
DQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN
YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR
MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVG
TALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL
ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESI
LPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKE
VLDATLIHQSITGLYETRIDLSQLGGDKRPAATKKAGQAKKKKLEGGGGSGSLPTCS
CLDRVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSH
GCPIAKWVLRRSSDEEKVLCLVRQRTGHHCPTAVMVVLIMVWDGIPLPMADRLYTE
LTENLKSYNGHPTDRRCTLNENRTCTCQGIDPETCGASFSFGCSWSMYFNGCKFGRS
PSPRRFRIDPSSPLHEKNLEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVARECRLG

SKEGRPFSGVTACLDFCAHP**H**R**D**IHNMNNGSTVVCTLTREDNRSLGVIPQDEQLHVL
PLYKLSDTDEFGSKEGMEAKIKSGAIEVLAPRRKKRTCFTQPVPRSGKKRAAMMTEV
LAHKIRAVEKKPIPRIKRKNNSTTTNNSKPSSLPTLGSNTETVQPEVKSETEPHFILKSS
DNTKTYSLMPSAPHPVKEASPGFSWSPKTASATPAPLKNDATASCGFSERSSTPHCTM
PSGRLSGANAAAADGPGISQLGEVAPLPTLSAPVMEPLINSEPSTGVTEPLTPHQPNHQ
PSFLTSPQDLASSPMEEDEQHSEADEPPSDEPLSDDPLSPAEEKLPHIDEYWSDSEHIFL
DANIGGVAIAPAHGSVLIECARRELHATTPVEHPNRNHPTRLSLVFYQHKNLNKPQH
GFELNKIKFEAKEAKNKKMKASEQKDQAANEGPEQSSEVNELNQIPSHKALTLTHDN
VVTVSPYALTHVAGPYNHWVGRPEFGSGEGRGSLLTCGDVEENPGP**VSKGEELFTG**
VVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTL
TYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGD
TLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTA
AGITLGMDELYKEFNSN***

3xFLAG ; SV40 NLS ; dSpCas9 ; nucleoplasmin NLS ; **Gly4Ser linker** ; **TET1 CD** (substitutions present in DED1 variant are indicated; **H → Y** and **D → A)** ; T2A self-cleaving peptide ; **eGFP** ; STOP (*****)

11. Curriculum Vitae

Vanja Tadić, research and teaching assistant, was born on 12th December 1990. She finished the undergraduate studies in Molecular Biology at Faculty of Science, University of Zagreb with a Bachelor's thesis "Alternative splicing in higher plants" in 2012. She finished the graduate studies in Molecular Biology at Faculty of Science, University of Zagreb with a Master's thesis "DNA replication in secondary lymphoid tissue" in 2014.

From 2014 she works as a research and teaching assistant at Faculty of Science, University of Zagreb in Zoldoš group, which focuses on Epigenetics. She works on a Croatian Science Foundation's project "Epigenetic regulation of IgG glycosylation". She is a co-author on two scientific papers. She was a part of several international scientific conferences and workshops with poster presentations. In 2016 she did a brief professional training for two months at Sanquin Research and Landsteiner Laboratory in Amsterdam.

She was a teaching assistant in Molecular Diagnostics, Medical Genetics, Molecular Genetics, Laboratory skill practice and Epigenetics. She was assistant supervisor of two Master's thesis at Faculty of Science, University of Zagreb and of one Master's thesis currently in development.